Adaptations in Speech Processing

DISSERTATION

zur Erlangung des akademischen Grades
Doctor rerum naturalium (Dr. rer. nat.) im Fach Psychologie

eingereicht an der
Lebenswissenschaftlichen Fakultät der Humboldt-Universität zu Berlin

von
Master in Literatur, Jue Xu

Präsidentin der Humboldt-Universität zu Berlin
Prof. Dr.-Ing. Dr. Sabine Kunst

Dekan der Lebenswissenschaftlichen Fakultät der Humboldt-Universität zu Berlin
Prof. Dr. rer. nat. Dr. rer. agr. Christian Ulrichs

Gutachter/innen

1. Prof. Dr. Werner Sommer

2. Prof. Dr. Reinhard Beyer

3. Prof. Dr. Birgit Stürmer

Datum der Einreichung: 19. Februar 2021

Tag der mündlichen Prüfung: 30. Juni 2021

Acknowledgements

The past few years were full of challenges. Now reaching the end of the journey I want to give my sincere thanks to many remarkable people who supported me in all respects throughout this time.

First of all, I wish to thank my supervisors, Prof. Dr. Werner Sommer and Prof. Dr. Rasha Abdel Rahman, for their constant encouragement and professional advice that have pushed me beyond my boundaries. My work and life have greatly benefitted from their reliable supports, their friendliness and their humanity. I am truly grateful to them.

Secondly, I would also like to thank all the colleagues from the neuro-cognitive psychology lab at Humboldt-Universität zu Berlin for such a friendly, supportive, and inspiring environment. I especially wish to thank our remarkable technician Guido Kiecker for always being there for us and providing the best technical support ever.

Last but not least, I am very grateful to my family, especially my parents for their constant support, trust, and health. Thank you!

Table of Contents

Zusammenfassung

Wie sich die Sprachwahrnehmung an ständig eingehende Informationen anpasst, ist eine Schlüsselfrage in der Gedanken- und Gehirnforschung. Die vorliegende Dissertation zielt darauf ab, zum Verständnis von Anpassungen an die Sprecheridentität und Sprachfehler während der Sprachverarbeitung beizutragen und unser Wissen über die Rolle der kognitiven Kontrolle bei der Sprachverarbeitung zu erweitern. Zu diesem Zweck wurden ereigniskorrelierte Potentiale (EKPs, englisch: event-related potentials, ERPs) N400 und P600 in der Elektroenzephalographie (EEG) analysiert. Die vorliegende Arbeit befasste sich insbesondere mit der Frage nach der Anpassung an die Sprecheridentität bei der Verarbeitung von zwei Arten von Sprachfehlern (Xu, Abdel Rahman, & Sommer, 2019), und untersuchte die proaktive Anpassungen, die durch die Erkennung von Sprachfehlern (Xu, Abdel Rahman, & Sommer, 2021) und durch die Sprecher(dis)kontinuität über aufeinanderfolgende Sätze in Situationen mit mehreren Sprechern ausgelöst wurden (Xu, Abdel Rahman, & Sommer, 2021, in press). Die Ergebnisse zeigten, dass unterschiedliche Sprachverarbeitungsstrategien entsprechend der Sprecheridentität von Muttersprachlern oder Nicht-Muttersprachlern und zwei verschiedenen Arten von Sprachfehlern angepasst wurden, was sich in unterschiedlichen N400- und P600-Effekten widerspiegelte. Darüber hinaus kann die Erkennung von Konflikten (Sprachfehler) und Sprecher(dis)kontinuität über aufeinanderfolgende Sätze hinweg eine proaktive kognitive Kontrolle erfordern, die die Verarbeitungsstrategien für den folgenden Satz schnell anpasst, was sich in bisher nicht gemeldeten sequentiellen Anpassungseffekten in der P600-Amplitude manifestierte. Basierend auf dem DMC Modell (Braver, 2012; Braver, Gray, & Burgess, 2007) und dem Überwachungsmodell der Sprachverarbeitung (van de Meerendonk, Indefrey, Chwilla, & Kolk, 2011) schlage ich vor, dass die P600-Amplitude nicht nur reaktive Anpassungen manifestiert, die durch Konflikterkennung ausgelöst werden, nämlich die klassischen P600-Effekte, die eine erneute Analyse der Sprachverarbeitung widerspiegeln, sondern auch proaktive Anpassungen in der Überwachung der Sprachverarbeitung, die Mechanismen der kognitiven Kontrolle von Aufmerksamkeit und Gedächtnis beinhalten.

Schlagwörter: Adaptation; Kognition; Erwartung; Erfahrung; N400; P600; Proaktive Kontrolle; Sequenzeffekt; Sprecheridentität; Sprachfehler

Abstract

How language perception adapts to constantly incoming information is a key question in mind and brain research. This doctoral thesis aims to contribute to the understanding of adaptation to speaker identity and speech error during speech processing, and to enhance our knowledge about the role of cognitive control in speech processing. For this purpose, event-related brain potentials (ERPs) N400 and P600 in the electroencephalography (EEG) were analyzed. Specifically, the present work addressed the question about adaptation to the speaker's identity in processing two types of speech errors (Xu, Abdel Rahman, & Sommer, 2019), and explored proactive adaptation initiated by the detection of speech errors (Xu, Abdel Rahman, & Sommer, 2021) and by speaker (dis-)continuity across consecutive sentences in multi-speaker situations (Xu, Abdel Rahman, & Sommer, 2021, in press). Results showed that different speech processing strategies were adapted according to native or non-native speaker identity and two different types of speech errors, reflected in different N400 and P600 effects. In addition, detection of conflict (speech error) and speaker (dis-)continuity across consecutive sentences engage cognitive control to rapidly adapt processing strategies for the following sentence, manifested in hitherto unreported sequential adaptation effects in the P600 amplitude. Based on the DMC model (Braver, 2012; Braver, Gray, & Burgess, 2007) and the monitoring theory of language perception (van de Meerendonk, Indefrey, Chwilla, & Kolk, 2011), I propose that the P600 amplitude manifests not only reactive adaptations triggered by conflict detection, i.e., the classic P600 effect, reflecting reanalysis of speech processing, but also proactive adaptations in monitoring the speech processing, engaging cognitive control mechanisms of attention and memory.

Keywords: adaptation; cognition; expectation; experience; N400; P600; proactive control; sequential effect; speaker identity; speech error

Synopsis

1 Introduction

1.1 Aims and Outline of the Present Work

In conversations involving multiple speakers, we are able to recognize and rapidly adapt to accents, the acoustic differences between speakers, varying proficiency or degrees of competence, and individual-specific characteristics in language use. The adaptation in speech processing is a sophisticated cognitive and linguistic tool, showing heightened sensitivity to social and linguistic context. However, there remain many unsolved puzzles and unexplored hypotheses based on existing findings. This doctoral thesis is dedicated to investigating the neurophysiological mechanisms of adaptation in speech processing by analyzing the event-related brain potentials (ERPs) in the electroencephalography (EEG).

A large body of EEG studies on speech processing investigates adaptation to social information such as the speaker's identity. Prior evidence suggests that neural correlates of speech processing may be modulated by speaker characteristics indicated by voices (Goslin, Duffy, & Floccia, 2012; van Berkum, van den Brink, Tesink, Kos, & Hagoort, 2008) and by native, foreign and regional accents (Grey & van Hell, 2017; Hanulíková, van Alphen, van Goch, & Weber, 2012; Romero-Rivas, Martin, & Costa, 2015). Yet most of these studies have presented listeners with foreign- or native-accented sentences only, devoid of any cues to the speaker's identity preceding the sentence. Hence, only after listeners recognized a female or child voice, or a non-native or regional accent as an indexical property of the speaker, could the processing of incoming signals begin to differ. It leaves no preparation time for the listener to form expectation or adjust their processing strategies in advance. Because individuals differ in their ability to recognize different accents, not providing cues to the speaker identity could lead to different ERPs in response to the errors (Grey & van Hell, 2017). Moreover, more recent studies have shown that neural correlates of speech processing can be affected by minimal visuo-social information such as a mere picture of the speaker's face (Grey, Cosgrove, & van Hell, 2020; Hernández-Gutiérrez et al., 2021). Face-to-face encounters in daily lives provide various social information before the talking even begins. Adaptations to speaker identity and to speaker (dis-)continuity in situations with multiple speakers may already happen during this preparation stage. Therefore, a presentation paradigm without preceding cues to the speaker identity is insufficient for the purpose of investigating adaptations to speaker identity in speech processing.

Furthermore, the investigation of adaptations in speech processing should consider the engagement of executive functions such as working memory, attention, and monitoring. Cognitive control or executive functions refer to the mechanisms of monitoring, regulating and guiding cognitive processes in sensory, memory and motor systems along internal goals (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Braver, 2012; Dreisbach & Fischer, 2012). Many behavioural and brain-imaging studies have shown that cognitive control is necessary for goal- and context-appropriate language processing and comprehension (see Blumstein, 2009; Key-DeLyria & Altman, 2016, for reviews). Yet it remains unclear how cognitive control is engaged in adaptation to context information in speech processing, for example, speaker identity and speech error detection. The ERP component P600, that was first characterized in the context of syntactic processing and has been extensively invoked for studying language processing, has been directly related to the P3b subcomponent of the P300 (Coulson, 1998; Coulson et al., 1998; Sassenhagen & Fiebach, 2019; Sassenhagen, Schlesewsky, & Bornkessel-Schlesewsky, 2014), which is typically invoked for studying cognitive control, for example, in the "oddball" and the task-switching paradigms (Jost, Mayr, & Rösler, 2008; see Leckey & Federmeier, 2019 for a review).

There remains a range of testable predictions for ERP studies regarding how cognitive control is engaged in strategic adaptation of speech processing to context information. One possible approach is to consider the conflict adaptation effect in cognitive control paradigms like Stroop, Flanker, and Simon tasks (see Gratton et al., 2018 for a review). Cross-domain conflict adaptation effects have already been reported cross the syntactic ambiguity resolution and the Stroop task (Hsu & Novick, 2016; Kan et al., 2013; Novick, Hussey, Teubner-Rhodes, Harbison, & Bunting, 2014; Thothathiri, Asaro, Hsu, & Novick, 2018; Vuong & Martin, 2014). Detecting a speech error in the previous sentence(s) may induce sustained control and affect how the upcoming sentence(s) is processed, which may be reflected in the P600. Another possible approach for exploring cognitive control in adaptation of speech processing is to consider the sequence of speakers cross consecutive sentences. In conversations involving multiple speakers, speakers naturally take their turns to speak. And listeners can tell who is going to talk from, for example, facial expression or body languages. Listeners may prepare for the upcoming speech when knowing there will be a switch or a repetition in the speaker's identity. In other words, the speaker (dis-)continuity across consecutive sentences might induce strategic adaptation in speech processing, which may be reflected in the P600. These two hypotheses about sequential adaptations triggered by the detection of speech errors in the preceding sentence and by the speaker (dis-)continuity across sentences have yet to be explored using EEG.

In conclusion, the overall aim of this doctoral thesis is to integrate the above-mentioned considerations and to explore how speech processing adapts to speaker identity and speech error, and how cognitive control is engaged in the process of adaptation in speech processing. Particularly, I asked the following questions:

1. How do error type and speaker nativeness modulate speech processing strategies?
2. Does detection of speech errors initiate proactive adaptation for the upcoming sentence?
3. Does speaker (dis-)continuity across consecutive sentences initiate proactive adaptation?
4. Does short-term experience with individual speakers with different error proneness persistently shape speech processing strategies for these speakers?

The answers to these questions are spread across three EEG studies conducted within this doctoral thesis. More specifically, through two EEG experiments in Study 1 I investigated the adaptation to native or non-native speaker identity in processing two types of speech errors, grammatical agreement violations and slips of the tongue (semantic blends in particular). In addition, I investigated how these adaptation effects might be influenced by tasks and error proportions in the experiment. In Study 2, with the purpose of exploring the role of cognitive control in speech processing, I re-analyzed the data from Experiment 2 in Study 1 to investigate sequential adaptation effects across consecutive sentences initiated by the detection of speech errors. Finally, to investigate adaptation to speaker identity in a yet unexplored manner, in Study 3 I examined sequential adaptation effects initiated by the speaker sequence (switch or repetition) across consecutive sentences. Additionally, I investigated whether or not short-term experience with individual speakers with different error proneness enduringly shifts individual-specific speech processing strategies.

In the following sections, I will introduce the theoretical background and the motivation for each study in more details. First, I will introduce two relevant ERP components and existing accounts in Section 1.2, followed by findings about accented speech processing in Section 1.3. Then I will present evidence about cognitive control engaged in speech processing, and introduce the conflict adaptation effects in cognitive control paradigms and across domains in Section 1.4. A dual mechanism model of proactive and reactive control

(DMC) will be introduced as a theoretical framework for the hypotheses (Braver, 2012; Braver, Gray, & Burgess, 2007). Finally, existing findings about adaptations in multi-speaker situations and the feedforward auditory streaming model (Lim, Shinn-Cunningham, & Perrachione, 2019; Lim, Tin, Qu, & Perrachione, 2019; Shinn-Cunningham, 2008) will be introduced to shape the hypotheses in Section 1.5. The three studies are subsequently summarized in Section 2, and jointly discussed in Section 3.

1.2 Electrophysiological Indicators of Speech Processing

The EEG is an electrophysiological monitoring method applied with the electrodes placed along the scalp to record electrical activity of the brain, which is primarily summed post-synaptic potentials of synchronously activated neurons in the neocortex. ERPs refer to stereotyped electrophysiological responses to a specific sensory, cognitive, or motor event. ERPs in the continuous online signal are widely used to inform conceptions of continuous internal processes, for example, language processing. Prior work identified two ERP components in the EEG correlated with processing semantic and syntactic information of speech: the N400 and the P600 component.

The N400 component is a negative voltage deflection peaking around 400 ms at centro-parietal sites, normally taken to reflect semantic processing and integration of verbal and non-verbal stimuli (Kutas & Federmeier, 2011; Kutas & Hillyard, 1980; van Berkum, 2004), but has also been taken to reflect prediction error (Rabovsky, Hansen, & McClelland, 2018). The P600 is a positive component maximal at centro-parietal sites starting around 500 ms after word onset, and may extend to one second or more. Initial P600 effects, i.e., increased P600 amplitudes, were observed in response to syntactic violations (Friederici, Pfeifer, & Hahne, 1993; Hagoort, Brown, & Groothusen, 1993; Osterhout & Holcomb, 1992). Therefore, the P600 has been suggested to reflect syntactic reanalysis or repair (e.g., Friederici et al., 1993). Later on, P600 effects were also seen in response to other kinds of linguistic deviations, without necessarily eliciting a preceding N400 effect (see Kuperberg, 2007, for a review), for example, locally ambiguous garden-path sentences (Kaan & Swaab, 2003; Osterhout & Holcomb, 1992), semantic reversal anomalies (van Herten, Kolk, & Chwilla, 2005), and orthographic errors (Vissers, Chwilla, & Kolk, 2006). Besides, P600 effects were also found to picture-sentence mismatches, in which sentences violated semantic information provided by preceding pictures (Vissers, Kolk, van de Meerendonk, & Chwilla, 2008). Therefore it has been proposed that the P600 does not just reflect syntactic processing but a more general reanalysis in speech perception (Kolk & Chwilla, 2007; Münte, Heinze, Matzke, Wieringa, & Johannes, 1998). The more recent Retrieval-Integration (RI) account of language processing suggested that the N400 amplitude reflects activation and retrieval of lexico-semantic information from long-term memory, and the P600 component indicates the resources required to integrate the activated (linguistic) information into a coherent mental representation of the utterance's content (Brouwer & Crocker, 2017; Brouwer, Crocker, Venhuizen, & Hoeks, 2017; Brouwer, Fitz, & Hoeks, 2012).

Many studies indicated the sensitivity of the P600 to the proportion or likelihood of violations within a given linguistic environment, being smaller under higher error proportions (e.g., Coulson, King, & Kutas, 1998; Hahne & Friederici, 1999). The P600 is also known to be task-sensitive, being larger in direct tasks like correctness judgments than indirect tasks like probe verifications (e.g., grammatical judgment vs. physical judgment in Gunter & Friederici, 1999; correctness judgment vs. semantic coherence judgment in Hahne & Friederici, 1998; correctness judgment vs. probe verification in Schacht, Sommer, Shmuilovich, Martienz, & Martin-Loeches, 2014). In contrast, the N400 is known to be only little affected by the experimental task and error proportion, taken to reflect relatively automatic processes in speech perception (Gunter & Friederici, 1999; Schacht et al., 2014). These findings about the sensitivity of P600 to domain-general factors of stimulus probability and task relevance indicate that language perception is not a purely automatic/unconscious process of structural processing but requires cognitive control, and the P600 manifests close interactions between language processing and cognitive control processes (e.g., monitoring, attention and working memory) (Coulson et al., 1998; Hahne & Friederici, 1999; Kolk & Chwilla, 2007). In line with this idea, the monitoring theory of language perception explains the P600 effects in terms of conflict monitoring, an important aspect of cognitive control (van de Meerendonk, Indefrey, Chwilla, & Kolk, 2011; van de Meerendonk, Kolk, Chwilla, & Vissers, 2009; Vissers et al., 2008). In typical cognitive conflict paradigms, for example the Simon (Simon, 1969) or the Flanker task (Gratton, Coles, & Donchin, 1992), conflicts occur between task-relevant and task-irrelevant stimulus properties and/or stimulus-response associations (see Gratton, Cooper, Fabiani, Carter, & Karayanidis, 2018, for a review). The monitoring theory of P600 suggests that during speech processing, conflicts may arise between listeners' expected linguistic input and what is actually encountered (auditory or written), triggering a reanalysis of the input for processing errors, reflected in P600 amplitude modulations.

Another domain-general interpretation, the P600-as-P3 account, considers the P600 as a variant of the P3b subcomponent of the P3, a domain-general brain response to salience (Coulson, 1998; Coulson et al., 1998; Sassenhagen & Fiebach, 2019; Sassenhagen et al., 2014). Accordingly, the P600 amplitude reflects the salience or significance of the stimulus category (Sassenhagen et al., 2014). The P600-as-P3 account was originally proposed based on the similarities between the stimuli and contexts that are known to elicit and/or affect the P600 and P3b components (Coulson, 1998; Coulson et al., 1998; see Leckey & Federmeier, 2019 for a review). The P3b component is normally elicited by uncertain, unexpected or surprising stimuli, and reflects the saliency of the stimuli, being highly sensitive to subjective aspects such as task demands, attention, global and local probability of the stimuli (Clayson & Larson, 2011; Coulson, 1998; Donchin, 1981; Gratton et al., 2018; Johnson, 1986). In comparison, the P600 is also elicited by surprising, incongruent, or intrusive stimuli, and is also sensitive to domain-general factors of stimulus probability, subjective salience and task relevance, showing similar electrophysiological properties and appears in similar contexts as the P3b (Sassenhagen & Fiebach, 2019; Sassenhagen et al., 2014).

The debate about the neural functions of P600 put aside, the overview of prior studies suggests that the P600 is a robust marker for understanding the online processing of language processes and how that processing changes with experience and context, which possibly engages cognitive control mechanisms. In the next section, I will elaborate findings and questions in the previous EEG research about how speaker identity, especially native or non-native speaker identity, affects speech processing.

1.3 Speaker Identity

It is well-established that speech processing device actively uses context information about a speaker's identity to anticipate upcoming speech. For example, stereotype-driven inferences about speaker characteristics such as sex, age or social status may trigger distinct N400 and P600 effects when perceiving incongruent versus congruent speech input (Goslin et al., 2012; Lattner & Friederici, 2003; van Berkum et al., 2008). Neural correlates of speech processing may also change in native, foreign and regional accents (Grey & van Hell, 2017; Hanulíková et al., 2012; Romero-Rivas et al., 2015). For example, Hanulíková and colleagues (2012) tested gender agreement violations and semantic world knowledge violations in native and Turkish-accented Dutch. They found a P600 effect to gender errors in native speech but not in non-native speech, whereas comparable N400 effects were elicited by semantic anomalies in both accents.

However, as explained above, the presentation paradigm without preceding cues about the speaker identity might be insufficient for the purpose of investigating adaptation to speaker identity. Because individuals differ in their ability to recognize different accents, and an adaptation to speaker identity may have already taken place during the preparation stage before the sentence was spoken. For example, Grey and van Hell (2017) found an N400-like effect to English subject pronoun errors only in a subset of listeners that correctly identified the foreign accent. More recent studies have shown that faces cueing speaker identity may modulate the neural correlates during speech comprehension (Grey et al., 2020; Hernández-Gutiérrez et al., 2021). Therefore, in experiments conducted within this doctoral thesis, faces were used as visual cues to provide explicit advance information about the upcoming speaker's identity. After seeing the speaker's face, spoken sentences were then presented, accompanied by the face picture. This should allow the listener to form and adapt their speech processing strategies ahead.

Furthermore, the error types most typically used in studying language processing are lexico-semantic and grammatical errors, both being atypical in the sense of daily encountering in conversations. In contrast, slips of the tongue, like Spoonerisms, such as "our queer old dean" rather than "our dear old queen", are more frequently encountered everyday speech errors. However, despite being of great interest for the study of speech production and comprehension, the neural correlates of perceiving slips of the tongue in native and non-native speech are yet to be investigated. Therefore, Study 1 used grammatical agreement violations – a typical type of speech error used in EEG studies, as well as (semantic and phonological) blends – a typical type of slips of the tongue. Depending on the native or non-native speaker identity, the processing of each error type should differ, reflected in distinct N400 and P600 effects.

Differences in accent and frequent errors typically distinguish non-native speech from native speech (Anderson-Hsieh, Johnson, & Koehler, 1992; Munro, 2003). Speech errors and especially grammatical errors are more frequent in non-native than native speech (Franceschina, 2005; Sabourin, Stowe, & De Haan, 2006). Based on knowledge about frequent or infrequent error types as a function of (native or non-native) speaker identity and/or based on acoustic features of non-native speech, neural correlates of processing non-native speech may differ from processing native speech. Therefore the existing findings that language processing is modulated by the nativeness of the speaker/speech lead to a further question: is it the acoustic features of non-native speech or rather the association of a higher error proneness with the speaker that has caused the processing of non-native speech to differ from native speech?

Evidence has been found that listeners take short-term experience about language use of individual speakers into consideration for sentence processing (Kroczek & Gunter, 2017; Regel, Coulson, & Gunter, 2010). Regel and colleagues (2010) presented participants with short passages of written text, ending with either literal or ironic statements made by one of two speakers in two sessions on two consecutive days, and manipulated the proportion of ironic statements made by the two speakers in each session. In Session 1, 70% ironic statements were made by the ironic speaker and 30% were made by the non-ironic speaker; in Session 2, it was 50% for each speaker. In Session 1, ironies of the non-ironic speaker elicited P600 effects relative to literal utterances, while ironies of the ironic speaker showed similar P600 amplitudes as literal utterances. In Session 2, P600 effects were found only for the (previously) ironic speaker but not for the non-ironic speaker. These results indicate that pragmatic knowledge about individual speakers can persistently affect language comprehension processes reflected in the P600 component (Regel et al., 2010). Therefore, considering the findings about absent P600 effects in non-native speech and under high error proportions, one may expect that short-term experience with individual speakers with different error proneness may enduringly shift individual-specific speech processing strategies. In Study 3, error proportions of three native speaker identities were manipulated to be differential in the first block of the experiment and changed into the same in the second block. P600 effects were analysed to assess whether short-term experience would shift individual-specific speech processing strategies in the second block.

To sum up, Study 1 and Study 3 were designed to investigate the questions about adaptations to speaker identity in speech processing, reflected in P600 modulations. Specifically, Study 1 focused on group-level adaptations to native or non-native speakers, whereas Study 3 focused on individual-level adaptations to speakers associated with varying error likelihood, independent of group-level stereotypical bias and acoustic influences.

1.4 Cognitive Control and Speech Processing

Cognitive or executive control must be exerted in many situations and tasks (Diamond, 2013) and may also be necessary for goal- and context-appropriate language processing and comprehension (see Key-DeLyria & Altman, 2016, for a review). Investigation about adaptation to speech error and speaker identity should consider possible engagement of cognitive control mechanisms in language processing, for example attention, monitoring, and working memory. However, this aspect has been underestimated in EEG studies about speech processing.

In cognitive control paradigms like Stroop, Flanker, and Simon tasks, the sequential effect, also known as the conflict adaptation effect, demonstrates that cognitive control processes are engaged continuously with finely tuned variations over very short time scales (see Gratton et al., 2018 for a review). For instance, participants tend to respond more quickly and more accurately to incongruent trials after incongruent rather than congruent trials (Hommel, Proctor, & Vu, 2004; Kerns et al., 2004; Rünger, Schwager, & Frensch, 2010; Stürmer, Leuthold, Soetens, Schröter, & Sommer, 2002). Of note, with few exceptions the expectancies governed by stimulus sequences are automatic and unconscious (Sommer, Leuthold, & Matt, 1998; Sommer, Matt, & Leuthold, 1990). Evidence has been found that conflict adaptation effects may transfer across tasks and domains. For example, January, Trueswell, and Thompson-Schill (2009) found overlapping BOLD activation to syntactic ambiguity and Stroop-like incongruency. Also sustained cognitive control initiated by previous conflicts (i.e., incongruent Stroop trials) is found to facilitate resolving syntactic ambiguities in subsequent sentences (Hsu & Novick, 2016; Novick et al., 2014; Thothathiri et al., 2018). Conversely, conflict detection in the syntactic domain seems to also facilitate conflict resolution in the Stroop task (Kan et al., 2013; Vuong & Martin, 2014). The findings about cross-domain conflict adaptation effects indicate either a domain-general cognitive control system that is shared across syntactic and non-syntactic domains or a domain-specific cognitive control system for syntactic and verbal conflicts. Processing syntactic ambiguity may initiate and be influenced by sustained cognitive control during sentence processing and comprehension. With this in mind, the sequential presentation paradigm used in EEG experiments is ideal for exploring sequential adaptation effects triggered by conflict detection – i.e., speech error, and speaker (dis-)continuity (explained in Section 1.5) in speech processing, which should greatly contribute to our understanding about how cognitive control is engaged in speech processing.

To explain conflict adaptation effects in cognitive conflict tasks, Braver, Gray, and Burgess (2007) proposed a dual mechanism model of proactive and reactive control (DMC). According to this model, if the processing of a stimulus produces a conflict, reactive control is triggered in order to resolve the conflict; in contrast, proactive control maintains task- or context-relevant information in memory and serves to anticipate and prevent conflict before it occurs (Braver, 2012; Braver et al., 2007). In language processing, the notion of cognitive control is related to the P600 by the monitoring theory of language perception (Sassenhagen et al., 2014; van de Meerendonk et al., 2011; van de Meerendonk et al., 2009; Vissers et al., 2008). Monitoring is an evaluative component in cognitive control that maintains contextual information active, and entails the detection of conflicts and the triggering of compensatory adjustments in control (Botvinick et al., 2001; Braver, 2012; Burgess & Braver, 2010). According to the monitoring theory of language perception, P600 reflects monitoring and conflict resolution of speech processing. A mismatch between competing linguistic representations or uncertainty in how to respond may be detected as conflict. And the P600 effect reflects a general reanalysis for processing errors, triggered by detecting such conflicts.

To conclude, considering the reports of cross-domain conflict adaptation effects and the DMC model, the monitoring theory would predict that conflicts in the syntactic domain in a given sentence should modulate proactive control over the processing of the following sentence, reflected in sequential adaptation effects in P600. In other words, previous-sentence error detection was expected to modulate the P600 in the current sentence. Therefore, Study 2 investigated whether the detection of syntactic errors in preceding sentences would affect P600 in correct and incorrect current sentences.

1.5 Multi-speaker Situations

A large body of ERP studies about the influence of speaker identity on speech processing used sentence-by-sentence presentations with multiple speakers in sequence. For each participant, one experiment session can be viewed as one novel linguistic and social situation. Much evidence has been found that in multi-speaker situations (i.e., situations with multiple speakers), participants show decrements in their performance measured in accuracy and speed as compared to single-speaker situations (situations with one speaker). For example, recognizing spoken words or phonemes in multi-speaker situations is slower and less accurate than in single-speaker situations (Green, Tomiak, & Kuhl, 1997; Kapadia & Perrachione, 2020; Nusbaum & Morin, 1992), even when speakers differ only slightly in pitch (Magnuson

& Nusbaum, 2007). Also increased neurophysiological responses have been found in multi- as compared to single-speaker situations (Chandrasekaran, Chan, & Wong, 2011; Kaganovich, Francis, & Melara, 2006; Wong, Nusbaum, & Small, 2004). With fMRI Wong and colleagues (2004) showed that word recognition in multi-speaker situations recruits not only traditional speech areas (e.g., posterior superior temporal cortex) but also areas associated with attention shifts (superior parietal cortex). In addition, recent studies demonstrated faster word identification when the speaker in the current trial is the same rather than a different person as in the previous trial (Carter, Lim, & Perrachione, 2019; Kapadia & Perrachione, 2020; Lim, Shinn-Cunningham, & Perrachione, 2019).

A feedforward auditory streaming model of speaker adaptation interprets these findings as evidence for listeners' adaptation to speaker (dis-)continuity in multi-speaker situations (Lim, Shinn-Cunningham, & Perrachione, 2019; Lim, Tin et al., 2019; Shinn-Cunningham, 2008). According to this model, the switch between multiple speakers across consecutive trials (speaker discontinuity) imposes attentional reorientation and interferes with working memory, whereas speaker repetition (continuity) facilitates speech processing in a feedforward manner. Speaker discontinuity disrupts the attentional focus during auditory streaming and interferes with working memory, inducing sequential and block-wise performance decrements. Considering the incremental encounter with the environment and the limited capacity of working memory, speaker (dis-)continuity (switch or repetition) across consecutive sentences in multi-speaker situations may affect the neural responses during speech processing. This hypothesis suggests an important role of cognitive control in adaptation to speaker identity of speech processing, which has yet to be investigated using EEG. Therefore, in Study 3 I assessed sequential effects triggered by speaker identity sequence (switch or repetition) to shed light on this specific question.

Together, Study 3 investigated two main research questions about adaptation to speaker identity in multi-speaker situations, (1) whether the speaker (dis-)continuity across consecutive sentences affects speech processing in a proactive manner, and (2) whether short-term experience with individual speakers with different error proneness induces enduring speaker-specific speech processing strategies, irrespective of acoustic features and stereotypical beliefs.

2 Summary of the Present Studies

2.1 Nativeness Effects in Speech Error Processing (Study 1)

The main goal of Study 1 was to provide evidence on how native or non-native speaker identities derived from advance visual cues and accents would affect speech processing of grammatical errors and slips of the tongue. Additionally, to understand the task sensitivity of P600 and to explore the neural correlates for perceiving slips of the tongue more thoroughly, two EEG experiments were included in Study 1.

In both experiments, German sentences that were either well-formed (correct) or contained a speech error (grammatical agreement violation or slip of the tongue), spoken in either native- or foreign-accented voices, were presented auditorily, randomly interspersed, preceded and accompanied by portraits of European or Asian faces, respectively. In Experiment 1, pictures of 90 European and 90 Chinese faces represented 180 speaker identities from two different ethnic backgrounds, whereas in Experiment 2 fifteen faces from each ethnic background, that is 30 speaker identities in total, were used. Each test session consisted of two blocks; each block included 180 trials, i.e., sentences. Experiment 1 and 2 were mainly differentiated in three aspects. First, Experiment 1 employed a probe verification task in 10% of all trials ($N = 36$), in which participants were instructed to judge whether a noun had been mentioned in the sentence or not; Experiment 2 employed a sentence correctness judgment task in all trials, in which participants judged the overall correctness of the sentence directly after hearing it. Second, the overall proportion of errors in the speech material was 66% in Experiment 1 and 50% in Experiment 2, and this was constant for both speaker conditions and error type conditions. Third, Experiment 2 included only semantic blends for the condition slips of the tongue in order to have a more homogeneous set of stimuli for this hitherto unexplored error type, whereas Experiment 1 also included phonological blends (20% phonological, 80% semantic blends in the material). In Experiment 1, 27 participants (16 female) were tested; in Experiment 2, 26 participants (20 female) were tested. All participants were native German speakers, gave informed consent, and received payment or course credits for participation. Continuous EEG was recorded from 64 Ag/AgCl electrodes arranged according to the extended 10/20 system during the whole test.

Experiment 1 revealed P600 effects elicited by grammatical agreement errors only in native speech in Block 2. In Experiment 2, grammatical errors evoked P600 effects in native but not in non-native speech, in line with prior studies (Hanulíková et al., 2012; Romero-Rivas et al., 2015); slips of the tongue in native speech elicited both N400 and P600 effects, while slips of the tongue in non-native speech engendered only P600 effects. Therefore, listeners seemed to rely more on a top-down processing strategy for non-native speech whereas more on bottom-up strategy for native speech, which will be discussed in details in Section 3.1. Besides, P600 effects were larger using a direct task in Experiment 2 than an indirect task in Experiment 1, while N400 effects were not influenced by the task. Moreover, in both experiments, short-term experience with speech errors resulted in more salient P600 effects in Block 2 relative to Block 1, irrespective of the accent.

In summary, Study 1 revealed that speech processing strategies were adapted according to native or non-native speaker identity dependent on the error type, namely, grammatical agreement violations and semantic blends, reflected in distinct N400 and P600 effects. Also, the results provided further evidence on the differential sensitivity to task, error proportion, and short-term experience of the N400 and P600 components.

2.2 Sequential Adaptation Effects in Speech Processing (Study 2)

Study 2 aimed to answer the question: can detection of speech errors initiate proactive control that influences the processing of the following sentence(s)? Therefore I reanalysed the data from Experiment 2 in Study 1 and assessed sequential adaptation effects in P600 after critical words across consecutive sentences.

The experimental task in Experiment 2, Study 1 was to judge the correctness of each sentence immediately after its presentation. Hence individual's correctness judgments could be used as subjective measures of whether an error was detected or not in previous sentences. Moreover, in the experiment, with equal probabilities, correct and incorrect sentences could be preceded by either a correct or incorrect sentence, enabling an investigation of correctness sequence effects across sentences. All data from Experiment 2, Study 1 were included and pre-processed in the same way. Four correctness sequences of two consecutive sentences were distinguished in the analyses: correct-incorrect (judged-as-correct previous sentence - incorrect current sentence), correct-correct, incorrect-incorrect, and incorrect-correct.

The results revealed a hitherto unreported sequential adaptation effect in sentence processing: P600 amplitudes to critical words in current sentences were smaller after detecting an error in the immediately preceding sentence than judging the preceding sentence as correct. This novel result was independent of and additive with the well-known immediate effect of syntactic sentence correctness on the P600. This finding indicated that the detection of speech errors initiates sustained proactive control over the monitoring demands for upcoming sentences. In other words, cognitive control can be triggered by the detection of speech errors and proactively affect the speech processing strategies.

2.3 Speaker (Dis-)continuity and Speaker-specific Error Proneness (Study 3)

Study 2 revealed a sequential adaptation effect of previous-sentence error detection on current-sentence P600, which was taken to reflect sustained proactive control over the monitoring demand for upcoming sentences. Considering the feedforward auditory streaming model of speaker adaptation (Lim, Shinn-Cunningham, & Perrachione, 2019; Lim, Tin et al., 2019; Shinn-Cunningham, 2008), similar sequential adaptation effects in P600 amplitudes were expected to be induced by the speaker (dis-)continuity. The first purpose of Study 3 was to assess this specific hypothesis. Furthermore, as mentioned in Section 1.3, motivated by the nativeness effects in Study 1 and by the previous research, the second research question of Study 3 was whether or not listeners adapt speech processing strategies for individual speakers according to the differential expectancy of speech errors based on short-term experience, independent of any acoustic features or stereotypical beliefs. Specifically, it was expected that short-term experience with different error proportions assigned to three speaker identities in the first half of the experiment would persistently modulate how listeners process speech (errors) from these speakers in the second half of the experiment where the error proportion was manipulated to be the same for all speakers.

In Study 3, a total of 39 native German-speaking participants (27 female) were tested. The same settings for electrophysiological recording and processing as in Study 1 were applied. Portraits of three Caucasian faces were used as visual cues and consistently assigned with three differential voices throughout the experiment to indicate the speaker identity. During the experiment, recorded sentences, which could be correct or contain a grammatical error, were presented sequentially, cued and accompanied by the portraits. Most importantly, each experiment was divided into two blocks. In Block 1, three native speakers (identified by face and voice) were associated with different error proportions (10%, 50%, and 90%). Participants cumulatively collected experience with speaker-specific error statistics in Block 1. In Block 2, the same three speakers were employed but all committed 50% incorrect sentences. Speaker-specific language processing and sequential effects of speaker identity in the P600 elicited by the critical words were analysed using ANOVA. Additionally, the ERP components N170 and N250r elicited by the face cues were analysed in order to provide complementary evidence that the speaker's identity associated with each face cue was learnt and recognized by the participants.

Results showed that P600 effects were elicited by speech errors in both blocks. Most importantly, regarding the first research question, in Block 1 where speakers had individualized error statistics, P600 amplitudes after critical words in current sentences (whether correct or not) were smaller when the speaker repeated as compared to when she switched between trials. Hence, when speakers differed in error proneness, listeners seemed to flexibly adapt their speech processing for each upcoming speaker.

Results regarding the second research question showed no speaker-specific differences in target word P600 effects in Block 2. Nevertheless, in Block 1, listeners showed higher accuracy in judging sentence correctness spoken by speakers with lower error proportions relative to speakers with higher error proportions: 10%-speaker > 50%-speaker > 90%-speaker accuracy; in Block 2 where all error proportions were 50%, there seemed to be no differences in the correctness judgment accuracies for the different speakers. Exploratory analyses on speaker-specific P600 in Block 1 were thus conducted, and indicated a trend that speech errors spoken by the three speakers elicited different P600 effects in Block 1. Additional analyses on P600 amplitude of the 50%-speaker in both blocks and enhanced P600 effects in Block 2 relative to Block 1 argued against a general attenuation of responses due to fatigue or practice over time. Most possibly, novel environmental statistics in Block 2 rapidly overwrote previous experience in Block 1, abolishing or reversing previous speaker-specific effects.

Additionally, when cue faces repeated across sentences, the N170 decreased and N250 increased relative to face changes, indicating adaptation and priming, respectively, of individual face identities, in line with the literature about face processing (Jacques & Rossion, 2006; Schweinberger, Pfütze & Sommer, 1995; see Schweinberger & Neumann, 2016, for a review). These face-specific effects indicate that speaker identities associated with the faces had been learned, processed and adapted to in both blocks, supporting our claim that the speaker sequence effect in Block 1 was indeed influenced by repetition or switch of the speaker identity between consecutive trials.

In conclusion, Study 3 revealed that in multi-speaker situations where speakers differ in error proneness, speaker discontinuity may impose attention reorientation and refreshment of speech processing strategies, whereas speaker continuity may proactively maintain neural resources for the repeated speaker to facilitate the upcoming speech processing. The result that no speaker-specific effects in Block 2 were found could be due to strategic control by the participants after realizing the altered error statistics in Block 2.

3 General Discussion

The present dissertation investigated adaptation to native and non-native speaker identities in processing two types of speech errors (Study 1), sequential adaptations initiated by the detection of speech errors (Study 2) and by the speaker (dis-)continuity across consecutive sentences (Study 3), and adaptations to short-term experience about error-proneness associated with individual speaker identities (Study 3) in speech processing. In the following sections, I will first offer a likely interpretation of the nativeness effects by suggesting different processing strategies for native and non-native speech. Questions about adapting to speaker-specific error proneness investigated in Study 3 will also be discussed in Section 3.1. In Section 3.2, I will discuss the role of cognitive control in speech processing with the focus on sequential adaptation effects found in Study 2 and 3. Finally, I will outline a proposal about the P600 indicating both reactive and proactive adaptations in speech processing based on all three studies in Section 3.3. Open questions and suggestions for further research will be discussed along the way.

3.1 Speaker Identity and Speech Error Processing

Together with the prior research, results in Study 1 suggested that listeners adapt different processing strategies for native and non-native speech, relying more on top-down strategies for processing non-native speech whereas more on bottom-up strategies for processing native speech. Above all, in line with the literature, P600 effects were sensitive to tasks and error proportions in Study 1. Therefore, P600 effects should not just reflect syntactic processing but rather a more general reanalysis in speech perception (Kolk & Chwilla, 2007; Münte et al., 1998), or as suggested by the Retrieval-Integration (RI) account, the resources required to integrate the activated information into a coherent mental representation of the utterance's content (Brouwer & Crocker, 2017; Brouwer et al., 2017). A more domain-general view of the P600 can also account for the results. The monitoring theory of language perception takes P600 effects to indicate a reanalysis of the input for processing errors triggered by detecting a conflict by the monitoring device (van de Meerendonk et al., 2011; van de Meerendonk et al., 2009; Vissers et al., 2008). A reanalysis of speech processing should involve necessary mechanisms with the purpose to resolve the encountered (linguistic) conflict, including linguistic mechanisms such as syntactic (re-)analysis and semantic integration of information, and a (re-)interpretation of the utterance (Sassenhagen & Fiebach, 2019). Moreover, the monitoring theory can explain the sequential adaptation effects found in

Study 2 and 3 better. P600 effects are thus interpreted as reflecting reanalysis of speech processing upon detection of conflicts (speech errors). In Section 3.3, I will combine findings from all three studies and specifically discuss the functional significance of the P600 and what it indexes in speech processing.

For grammatical agreement violations, a reanalysis of speech processing was only triggered in native speech, not in non-native speech. Non-native accents present acoustic and prosodic complexities for speech processing, making it harder for the listener to correctly recognize words in a bottom-up way (Anderson-Hsieh et al., 1992; Munro, 2003). Additionally, based on daily experience and/or stereotypical beliefs, non-native speakers have difficulties meeting grammatical agreements in natural speech (Franceschina, 2005; Sabourin et al., 2006). Moreover, this type of speech errors are actually errors in word forms realized in inflectional morphemes, which don't necessarily hinder retrieving and apprehending the core meaning of the utterance. Hence the non-native accent and the expectation of word form errors may render the non-native speech seeming less suitable for a bottom-up strategy based on word form information. So listeners adapt a strategy that actively suppresses processing word forms and concentrates on interpreting the approximate meaning and intention of the utterance for an upcoming non-native speaker. Therefore reanalysis processes after grammatical errors were observed only in native but not in non-native speech.

Results regarding the N400 effect engendered by the other type of error, semantic blends (slips of the tongue), also supported this idea that listeners relied less on bottom-up strategies when perceiving non-native speech: the activation and retrieval of lexico-semantic information was reduced for semantic blends in non-native relative to native speech. In comparison, naïve semantic violations typically used to study ERPs in speech processing, namely, word substitutions, normally induce N400 effects regardless of the accent and P600 effects only in native speech (e.g., Romero-Rivas et al., 2015). In other words, in native speech semantic blends were processed in a similar way as semantic violations (i.e., with an N400 and a P600 effect), whereas in non-native speech blends induced different effects compared to semantic violations (i.e., only an N400 effect but no P600 effect) (Romero-Rivas et al., 2015). Semantic violations are salient anomalies in their phonological forms, whereas semantic blends consist of fragments of the legal words/phrases that make sense in the linguistic context, highly resembling the intended words in word form and pronunciation. The subtle differences in word forms seem to be suppressed or ignored in processing non-native speech, since they do not directly hinder the sentence interpretation. Therefore, listeners rely more on top-down processing for non-native speech, which directs less attention to word forms and makes less effort to retrieve lexico-semantic information; whereas in native speech listeners rely more on bottom-up processing strategies. Thus semantic blends in native speech induced lexico-semantic retrieval as well as reanalysis processes, similar as typical semantic violations.

Notably, different from grammatical violations, semantic blends induced reanalysis processes in non-native as well as in native speech. It is possibly attributed to the varying degrees of expectancy for the error types with regard to native or non-native speaker identity. Higher expectation results in lower saliency of encountered conflicts (speech errors), reflected in the P600 effect. Compared to grammatical agreement errors that were expected because of the non-native identity, semantic blends were much less associated with any particular speaker type and, thus, elicited similar P600 effects in native and non-native speech. Future studies can examine whether the present results can be generalized to other categories of speech errors regarding the expectancy associated with the speaker identity, especially slips of the tongue; also, depending on the locus of failure within the speech production process, there might be differences in the perception of different types of slips of the tongue.

Based on Study 1 and the previous research, Study 3 was designed to assess whether short-term experience with speaker-specific error proneness would persistently shift processing strategies, but no individual-specific effects were found in Block 2, where the error proportion was manipulated to be the same for the speakers. However, this result does not necessarily offer evidence against speaker-specific processing strategies in Block 1, where the speakers had different error proportions. For example, Regel, Coulson, and Gunter (2010) assessed speaker-specific P600 effects elicited by ironic statements relative to literal utterances by manipulating irony proportions between two speakers, differential in Session 1 and equal in Session 2. Their result indicated that novel environmental statistics in a new situation rapidly overwrote previous experience and reversed previous effects. In fact, exploratory analyses on Block-1 P600 in Study 3 indicated a trend that speech errors spoken by the three speakers elicited different P600 effects. In addition, analyses on accuracy in sentence correctness judgments revealed a negative correlation between accuracy and speaker-specific error proportion in Block 1. Therefore, it is likely that listeners adapted speaker-specific speech processing strategies in Block 1, but they noticed the new equally distributed error statistics across the three speakers in Block 2 and quickly adapted and employed the same or similar speech processing strategies for these speakers. In comparison, the belief of a higher error proneness for non-native speakers is stereotypically associated and/or experience-based, thus is more likely to persistently affect speech processing strategies regardless of the actual environment error statistics during the experiment. How speaker-specific error proneness affects speech processing regardless of acoustic features and stereotypical beliefs should be further clarified in future studies, for example, by optimizing the experiment design of Study 3 through altering the block sequence, adjusting the error proportions in both blocks, or manipulating number and variety of face cues.

3.2 Sequential Adaptations in Speech Processing

In light of the assumptions from the literature that (1) a variety of cognitive functions are used during sentence processing and comprehension, (2) processing syntactic ambiguity may initiate and be influenced by sustained cognitive control reflected in sequential effects in accuracy and reaction speed (Hsu & Novick, 2016; Kan et al., 2013; Novick et al., 2014; Thothathiri et al., 2018; Vuong & Martin, 2014), and (3) listeners show performance decrements and increased neural activities in multi-speaker situations as compared to single-speaker situations (Chandrasekaran et al., 2011; Green et al., 1997; Kaganovich et al., 2006; Kapadia & Perrachione, 2020; Magnuson & Nusbaum, 2007; Nusbaum & Morin, 1992; Wong et al., 2004), Study 2 and 3 intended to explore sequential adaptations in speech processing. Hitherto unreported sequential effects were indeed found in P600 amplitudes in the current sentence triggered by error detection in the preceding sentence (Study 2) and by the speaker (dis-)continuity across consecutive sentences (Study 3).

Experimental sessions build specific local environments. According to the context updating model (Donchin, 1981; Donchin & Coles, 1988), participants' expectations are governed by their contextual models of the situation, which are continuously updated. From the perspective of the participant, default expectation is for sentences to be well-formed, but encountering errors might change this expectation by updating the contextual model to the local environment. Considering the incremental encounter with the environment and the limited capacity of working memory, sequential adaptation in sentence processing is thus to be expected for the sake of updating one's contextual model of the experiment situation. Study 2 went beyond the previous research by analysing sequential effects in speech processing and found that P600 amplitudes to critical words in current sentences (correct and incorrect) were smaller after detecting an error in the immediately preceding sentence than judging the preceding sentence as correct. This sequential effect in P600 is related to a rapid adaptation in speech processing as a result of updating the mental model of the environment according to the context set up by the preceding sentence(s). Encountering a linguistic error leads to a higher expectancy of linguistic conflicts or complexities in the upcoming speech input, which may result in heightened attention for conflicts and proactive maintenance of relevant neural resources for the upcoming sentence. Given that there is little or no syntactic conflict in correct current sentences, the decreased P600 amplitudes in these sentences are most naturally attributed to a shared general mechanism rather than error-related mechanisms, most likely the deployment of resources for monitoring speech processing, which entails attention control and memory resources, as suggested in the monitoring theory of language perception (van de Meerendonk et al., 2009; van de Meerendonk et al., 2011; Vissers et al., 2008). When error expectation is increased by error detection in the preceding sentence, the relative saliency of words in the next sentence becomes lower. After detecting a speech error in the previous sentence, the neural resources for monitoring the speech processing are proactively maintained, resulting in smaller P600 amplitudes in the next sentence.

Based on findings in Study 1 and 2 and the previous research about adaptation in multi-speaker situations, Study 3 further asked whether the speaker (dis-)continuity (i.e., sequence, switch or repetition) across consecutive sentences in a multi-speaker situation would trigger sequential adaptation in speech processing. Increased P600 amplitudes were indeed found after critical words in current sentences (whether correct or not) when the speaker switched as compared to repeated between sentences in situations where speakers differed in error proneness. Accordingly, face ERPs N170 and N250r analyses showed typical adaptation and priming effects after a repetition of the same face identity relative to a switch of the face identity, supporting the claim that speaker (dis-)continuity triggered the sequential effects in P600 in Block 1. In line with the feedforward auditory streaming model of speaker adaptation (Lim, Shinn-Cunningham, & Perrachione, 2019; Lim, Tin et al., 2019; Shinn-Cunningham, 2008), listeners rapidly adapted to an upcoming, i.e., about-to-speak, speaker identity in multi-speaker situations, reflected by sequential effects triggered by the speaker (dis-)continuity. More specifically, speaker discontinuity disrupts the attentional focus during auditory speech perception, imposes attention reorientation and “reset” speech processing strategies, whereas speaker continuity proactively maintains neural resources for the repeated speaker active, hence facilitating the upcoming speech processing in a feedforward manner. In other words, a switch of the speaker’s identity (speaker discontinuity) may “reset” the speech processing strategies, that is, reactivate or boost monitoring, hence increasing P600 amplitudes; a repetition of the speaker’s identity (speaker continuity) may “prime” previously activated resources for speech processing, meaning that fewer resources would have to be additionally recruited for a repeated speaker, hence decreasing P600 amplitudes. Sequential adaptation of P600 was found only in Block 1 where speakers differed in error proportions, but not in Block 2 where all speakers changed into equal error proportions. It seems that listeners only adapted to the speaker identity from trial to trial if necessary. As suggested in Section 3.1, listeners possibly quickly adapted to new environment statistics in Block 2, resulting in similar strategies for all speakers in Block 2, making it no longer necessary for sequentially adapting to speaker (dis-)continuity.

In the next section, I will combine findings from all three studies and propose that P600 manifests reactive and proactive adaptations in speech processing under the theoretical framework of the DMC model (Braver, 2012; Braver et al., 2007) and the monitoring theory of language perception (van de Meerendonk et al., 2009; van de Meerendonk et al., 2011; Vissers et al., 2008).

3.3 The P600 as Indicator of Reactive and Proactive Adaptations

Together the present studies have revealed that speech processing strategies are adapted to (native or non-native) speaker identity and error type (Study 1), detection of conflict (speech error) (Study 2) and speaker (dis-)continuity across consecutive sentences (Study 3). The sensitivity of P600 effects to error proportion, task, and short-term experience in Study 1, sequential effects in Study 2 and 3, as well as a number of prior studies (e.g., Coulson et al., 1998; Gunter & Friederici, 1999; Hahne & Friederici, 1998, 1999; Schacht et al., 2014), consistently revealed a close interaction between cognitive control and P600 amplitudes. These results do not fit with the notions that interpret P600 effects only as syntactic or linguistic processing. Instead, the current findings about “trial-to-trial” adaptations reconcile with the monitoring theory of language perception (van de Meerendonk et al., 2011; van de Meerendonk et al., 2009; Vissers et al., 2008) and even the P600-as-P3 account (Coulson, 1998; Coulson et al., 1998; Sassenhagen & Fiebach, 2019; Sassenhagen et al., 2014).

The monitoring theory of language perception (van de Meerendonk et al., 2011; van de Meerendonk et al., 2009; Vissers et al., 2008) was based on the notion about conflict monitoring, detection, and adjustment in typical cognitive control paradigms (Botvinick et al., 2001; Botvinick, Cohen & Carter, 2004; Kerns et al., 2004). Speech processing engages ongoing monitoring for conflict such as linguistic errors and complexities; once conflict is detected, compensatory mechanisms (reanalysis) are recruited to resolve the conflict, adjust expectancy and processing strategies for future events. Incoming information is dynamically updated through increased engagement of attentional control mechanisms to adapt more efficient strategies for perception and goal-appropriate responding (i.e., tasks). The dual mechanisms model of cognitive control (DMC, Braver, 2012; Braver et al., 2007) is closely related to the monitoring theory (Botvinick et al., 2001; Botvinick et al., 2004; Kerns et al., 2004). The DMC model focuses on the timescale of implementation of cognitive control in typical cognitive control paradigms, making a distinction between proactive and reactive control processes. Reactive control processes are activated after the occurrence of an imperative stimulus in a transient, context-dependent manner, similar as the reanalysis mechanism suggested in the monitoring theory. In contrast, proactive control maintains task- or context-relevant information in memory and serves to anticipate and prevent conflict before it occurs, similar as the monitoring mechanism suggested in the monitoring theory.

In line with the monitoring theory of language perception (van de Meerendonk et al., 2009; van de Meerendonk et al., 2011; Vissers et al., 2008) and based on the DMC model (Braver, 2012; Braver et al., 2007), I propose that P600 amplitude manifests not only the reactive reanalysis of speech processing, i.e., the resolution of linguistic conflicts (the classic P600 effect), but also proactive speech monitoring involving cognitive control mechanisms. Similar as in the DMC, speech processing is adapted proactively and reactively, distinguished by the time point of establishment of adaptation. Reactive adaptations are triggered reactively by detecting a conflict (i.e., linguistic errors or complexities) during online speech processing (listening or reading). In contrast, proactive adaptations are triggered and established in advance, for instance, in the preceding sentence or when seeing the speaker before the sentence begins to unfold. In this notion, classic P600 effects, namely, increased P600 amplitudes in response to linguistic errors or uncertainty in processing, are taken to reflect reanalysis of speech processing, which involves necessary neural mechanisms with the purpose to resolve the encountered (linguistic) conflict or uncertainty in processing, for example, syntactic (re-)analysis, semantic integration of information, and a (re-)interpretation of the utterance (Sassenhagen & Fiebach, 2019). Also, the size of P600 effects reflects the saliency of encountered conflicts, which depends on the type of error with regard to the proactively established expectation based on, for instance, the speaker identity (Study 1). Thus, depending on the encountered error type, the relative saliency and the reanalysis may vary. Of course this does not mean explicit discrimination of the linguistic category of speech errors, but rather refers to which aspects of linguistic processing does the encountered error hinder during online speech processing. In short, P600 amplitudes manifest reactive conflict resolution triggered by error detection – that is the classic P600 effect, reflecting reanalysis of speech processing, as well as reactive adaptation to error type with regard to existing proactively-adapted expectations and strategies.

On the other hand, P600 amplitudes also manifest the proactive deployment of resources for monitoring speech processing, engaging attention and memory resources. The proactive adaptation involves a sustained change in attentional weights when preparing to undertake a task – proactive maintenance of goal-relevant information (Braver, 2012), as well as trial-by-trial shifts in attentional bias and/or response threshold in anticipation of conflicts, for instance, when using cue (speaker) information to prepare for an upcoming sentence. The first type of proactive adaptation is established based on environment statistics and task demands. P600 effects were enhanced under lower error proportion and in a direct compared to an indirect task. Hence the reanalysis of speech processing and the saliency of encountered conflicts, reflected in the P600 effect, depend strongly on where the attention is directed to within a specific environment. The experience-based information about task and error proportion is sustainably maintained and updated in the context model in short-term memory, proactively affecting resource allocation for conflict resolution and monitoring. Furthermore, knowledge about the upcoming speaker identity can also be considered as a kind of information that proactively affects how the upcoming sentence is processed, resulting in nativeness effects in the P600 effect in Study 1. The second type of proactive adaptation is revealed by sequential effects in Study 2 and 3. The neural resources for monitoring speech processing is continuously “fine-tuned” over very short time scales depending on conflict detection in the preceding sentence and speaker (dis-)continuity across consecutive sentences. Also the necessity of a constantly-adapted strategy depends on the situation (context), whether there are differences between the speakers’ language competence/use in a multi-speaker situation or not (Study 3).

Although Study 2 found additive effects of previous-trial error detection on current-trial conflict resolution (P600 effects), it does not prove the independence of cue/face-induced adaptation and the sequential adaptation in speech processing. The cue/face-induced adaptation (P600 effects) and sequential adaptation across sentences were not directly contrasted in the present experiment design. The monitoring device should feedback critical events such as error detection, speaker identity, and speaker (dis-)continuity to trigger reactive conflict resolution as well as proactive strategic adaptation, and the proactive strategic control may at least affect the monitoring status. It remains debatable whether sequential adaptations across sentences share the same sources as the proactive adaptation in conflict resolution triggered by current-trial cue (face). It is possible that distinguishable types of cognitive control are triggered by current-sentence cues and previous-sentence conflicts. Future studies should investigate whether both effects recruit the same or similar strategic control processes, for example by testing whether current-sentence speaker identity can override effects of sequential adaptations that are previously triggered (e.g., Alpay, Goerke, & Stürmer, 2009).

The current finding about sequential effects in the P600 seems to reconcile with the P600-as-P3 account (Coulson et al, 1998; Leckey & Federmeier, 2019), especially that the P600 showed similar patterns of trial-to-trial modulations triggered by speaker (dis-)continuity as the P3b component in task-switching paradigms: P3b amplitudes were larger for task switch trials relative to repeat trials (Kopp, Steinke, & Visalli, 2020; see Kiesel et al., 2010 for a review). Also, in contrast to the random task-switching blocks, in single-task blocks, in which trial-by-trial updating is not necessary, no sequential P3b was elicited by the cues before the targets (Jost et al., 2008). The current finding makes a range of testable predictions for future studies regarding the relationship between P3b and P600. For example, future studies can use the current presentation paradigm with face cues preceding sentences and combine prior findings in task-switching or -cueing paradigms regarding cognitive control into investigating language processing and speaker identity.

3.4 Conclusions

In conclusion, the present work explored how speech processing adapts to speaker identity and how cognitive control is engaged in the process of adaptation in speech processing. The presentation paradigm with face cues preceding sentences allowed the listener to form expectations based on speaker identity before the sentence began to unfold, hence enabled the investigation of sequential adaptations in sentence processing with regard to speaker identity.

The results suggest that speech processing strategies are adapted to native or non-native speaker identity and different types of speech errors, and that detection of conflict and speaker (dis-)continuity engage proactive control to rapidly adapt strategies for monitoring speech processing in the following sentence. Based on the DMC model (Braver, 2012; Braver et al., 2007) and the monitoring theory of language perception (van de Meerendonk et al., 2011), these results are taken to suggest that P600 amplitudes reflect reactive and proactive adaptations in speech processing.

The novel findings of sequential adaptation effects in P600 emphasized a perhaps underestimated important role of cognitive mechanisms in speech processing. Analysis of sequential effects across consecutive sentences, as used in the present study, provides a valuable tool for investigating cognitive control in language processing and the functions of the P600-reflected neural activities. This idea makes a range of testable predictions for future studies regarding the larger question about the domain specificity of the processes used to comprehend (and produce) language. Neurocognitive models of language should allow for a general executive component (such as the monitoring device) operating during or at least in support of linguistic parsing. Future studies can assess the scope of proactive adaptation processes, for example, as to the role of the encountered linguistic problems (e.g., sequential effects in semantic processing), the lengths of the sequential effects, or the role of specific speaker identity, and seek evidence how proactive and reactive processes might interact in language processing. It also remains to be investigated whether engagement of proactive control is required for language processing during more communicative scenarios – such as during passive reading or using comprehension questions as tasks – or for adaptation in linguistic contexts such as stories.

References

Alpay, G., Goerke, M., & Stürmer, B. (2009). Precueing imminent conflict does not override sequence-dependent interference adaptation. *Psychological Research PRPF*, *73*(6), 803-816. doi: 10.1007/s00426-008-0196-9

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentais, prosody, and syllable structure. *Language learning, 42*(4), 529-555

Blumstein, S. (2009). Reflections on the Cognitive Neuroscience of Language. In Gazzaniga MS (Ed.), *The Cognitive Neurosciences*. Cambridge, US: MIT Press

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological review, 108*(3), 624. doi: 10.1030//0033-295X.108.3.624

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: an update. *Trends in cognitive sciences*, *8*(12), 539-546. doi: 10.1016/j.tics.2004.10.003

Braver, T. S. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends in cognitive sciences, 16*(2), 106-113. doi: 10.1016/j.tics.2011.12.010

Braver, T. S., Gray, J. R., & Burgess, G. C. (2007). Explaining the many varieties of working memory variation: Dual mechanisms of cognitive control. In A. R. A. Conway, C. Jarrold, M. J. Kane (Eds.) & A. Miyake & J. N. Towse (Ed.), *Variation in working memory* (pp. 76-106). New York, US: Oxford University Press

Brouwer, H., Crocker, M. W., Venhuizen, N. J., & Hoeks, J. C. (2017). A neurocomputational model of the N400 and the P600 in language processing. *Cognitive science*, *41*, 1318-1352. doi: 10.1111/cogs.12461

Brouwer, H., Crocker, M. W., Venhuizen, N. J., & Hoeks, J. C. (2017). A Neurocomputational Model of the N400 and the P600 in Language Processing. *Cognitive science, 41*(S6), 1318-1352. doi: 10.1111/cogs.12461

Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about semantic illusions: rethinking the functional role of the P600 in language comprehension. *Brain research*, *1446*, 127-143. doi: 10.1016/j.brainres.2012.01.055

Burgess, G. C., & Braver, T. S. (2010). Neural mechanisms of interference control in working memory: effects of interference expectancy and fluid intelligence. *PLoS One, 5*(9), e12861. doi: 10.1371/journal.pone.0012861

Carter, Y. D., Lim, S.-J., & Perrachione, T. K. (2019). *Talker continuity facilitates speech processing independent of listeners' expectations.* Paper presented at the 19th International Congress of Phonetic Sciences. retrieved from https://sites.bu.edu/cnrlab/files/2019/05/Carter-Lim-Perrachione-2019-ICPhS-ERTA.pdf

Chandrasekaran, B., Chan, A. H., & Wong, P. C. (2011). Neural processing of what and who information in speech. *Journal of Cognitive Neuroscience, 23*(10), 2690-2700. doi: 10.1162/jocn.2011.21631

Clayson, P. E., & Larson, M. J. (2011). Conflict adaptation and sequential trial effects: Support for the conflict monitoring theory. *Neuropsychologia*, *49*(7), 1953-1961. doi: 10.1016/j.neuropsychologia.2011.03.023

Coulson, S. (1998). ERPs and domain specificity: Beating a straw horse. *Language and cognitive processes, 13*(6), 653-672. doi: 10.1080/016909698386410

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and cognitive processes, 13*(1), 21-58. doi: 10.1080/016909698386582

Diamond, A. (2013). Executive Functions. *Annual Review of Psychology, 64*, 135-168. doi: 10.1146/annurev-psych-113011-143750

Donchin, E., & Coles, M. G. (1988). Is the P300 component a manifestation of context updating. *Behavioral and brain sciences*, *11*(3), 357-427

Donchin, E. (1981). Surprise!… surprise?. *Psychophysiology*, *18*(5), 493-513

Dreisbach, G., & Fischer, R. (2012). Conflicts as aversive signals. *Brain and cognition, 78*(2), 94-98. doi: 10.1016/j.bandc.2011.12.003

Franceschina, F. (2005). *Fossilized second language grammars: The acquisition of grammatical gender* (Vol. 38): John Benjamins Publishing

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive Brain Research, 1*(3), 183-192. doi: 10.1016/0926-6410(93)90026-2

Goslin, J., Duffy, H., & Floccia, C. (2012). An ERP investigation of regional and foreign accent processing. *Brain Lang, 122*(2), 92-102. doi:10.1016/j.bandl.2012.04.017

Gratton, G., Coles, M. G., & Donchin, E. (1992). Optimizing the use of information: strategic control of activation of responses. *Journal of Experimental Psychology: General, 121*(4), 480. doi: 10.1037/0096-3445.121.4.480

Gratton, G., Cooper, P., Fabiani, M., Carter, C. S., & Karayanidis, F. (2018). Dynamics of cognitive control: Theoretical bases, paradigms, and a view for the future. *Psychophysiology, 55*(3), e13016. doi: 10.1111/psyp.13016

Green, K. P., Tomiak, G. R., & Kuhl, P. K. (1997). The encoding of rate and talker information during phonetic perception. *Perception & Psychophysics, 59*(5), 675-692. doi: 10.3758/BF03206015

Grey, S., Cosgrove, A. L., & van Hell, J. G. (2020). Faces with foreign accents: An event-related potential study of accented sentence comprehension. *Neuropsychologia, 147*, 107575.doi: 10.1016/j.neuropsychologia.2020.107575

Grey, S., & van Hell, J. G. (2017). Foreign-accented speaker identity affects neural correlates of language comprehension. *Journal of Neurolinguistics, 42*, 93-108. doi:10.1016/j.jneuroling.2016.12.001

Gunter, T. C., & Friederici, A. D. (1999). Concerning the automaticity of syntactic processing. *Psychophysiology, 36*(1), 126-137. doi: 10.1017/S004857729997155X

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and cognitive processes, 8*(4), 439-483. doi: 10.1080/01690969308407585

Hahne, A., & Friederici, A. D. (1998). ERP-evidence for autonomous first-pass parsing processes in auditory language comprehension. *Journal of Cognitive Neuroscience, Supplement, 125*

Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. *Journal of Cognitive Neuroscience, 11*(2), 194-205. doi: 10.1162/089892999563328

Hanulíková, A., van Alphen, P. M., van Goch, M. M., & Weber, A. (2012). When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience, 24*(4), 878-887. doi: 10.1162/jocn_a_00103

Hernández-Gutiérrez, D., Muñoz, F., Sánchez-García, J., Sommer, W., Abdel Rahman, R., Casado, P., Jiménez-Ortega, L., Espuny J., Fondevila S., & Martín-Loeches, M. (2021). Situating Language in a Minimal Social Context: How Seeing a Picture of the Speaker's Face Affects Language Comprehension. *Social Cognitive and Affective Neuroscience*. doi: 10.1093/scan/nsab009

Hommel, B., Proctor, R. W., & Vu, K.-P. L. (2004). A feature-integration account of sequential effects in the Simon task. *Psychological Research, 68*(1), 1-17. doi: 10.1007/s00426-003-0132-y

Hsu, N. S., & Novick, J. M. (2016). Dynamic engagement of cognitive control modulates recovery from misinterpretation during real-time language processing. *Psychological science, 27*(4), 572-582. doi: 10.1177/0956797615625223

Jacques, C., & Rossion, B. (2006). The speed of individual face categorization. Psychological science, 17(6), 485-492. doi: 10.1111/j.1467-9280.2006.01733.x

Johnson, R. (1986). A triarchic model of P300 amplitude. *Psychophysiology*

Jost, K., Mayr, U., & Rösler, F. (2008). Is task switching nothing but cue priming? Evidence from ERPs. *Cognitive, Affective, & Behavioral Neuroscience*, *8*(1), 74-84. doi: 10.3758/CABN.8.1.74

Kaan, E., & Swaab, T. Y. (2003). Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of Cognitive Neuroscience, 15*(1), 98-110. doi: 10.1162/089892903321107855

Kaganovich, N., Francis, A. L., & Melara, R. D. (2006). Electrophysiological evidence for early interaction between talker and linguistic information during speech perception. *Brain research, 1114*(1), 161-172. doi: 10.1016/j.brainres.2006.07.049

Kan, I. P., Teubner-Rhodes, S., Drummey, A. B., Nutile, L., Krupa, L., & Novick, J. M. (2013). To adapt or not to adapt: The question of domain-general cognitive control. *Cognition, 129*(3), 637-651. doi: 10.1016/j.cognition.2013.09.001

Kapadia, A. M., & Perrachione, T. K. (2020). Selecting among competing models of talker adaptation: Attention, cognition, and memory in speech processing efficiency. *Cognition, 204*, 104393. doi: 10.1016/j.cognition.2020.104393

Kerns, J. G., Cohen, J. D., MacDonald, A. W., Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science, 303*(5660), 1023-1026. doi: 10.1126/science.1089910

Key-DeLyria, S. E., & Altmann, L. J. (2016). Executive function and ambiguous sentence comprehension. *American journal of speech-language pathology, 25*(2), 252-267. doi: 10.1044/2015_AJSLP-14-0111

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching — A review. *Psychological bulletin*, *136*(5), 849. doi: 10.1037/a0019842

Kolk, H., & Chwilla, D. (2007). Late positivities in unusual situations. *Brain and language, 100*(3), 257-261. doi:10.1016/j.bandl.2006.07.006

Kopp, B., Steinke, A., & Visalli, A. (2020). Cognitive flexibility and N2/P3 event-related brain potentials. *Scientific reports*, *10*(1), 1-17.doi: 10.1038/s41598-020-66781-5

Kroczek, L. O. H., & Gunter, T. C. (2017). Communicative predictions can overrule linguistic priors. *Sci Rep, 7*(1), 17581. doi:10.1038/s41598-017-17907-9

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain research, 1146*, 23-49

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual review of psychology, 62*, 621-647

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*(4427), 203-205. doi: 10.1126/science.7350657

Lattner, S., & Friederici, A. D. (2003). Talker's voice and gender stereotype in human auditory sentence processing–evidence from event-related brain potentials. *Neuroscience letters, 339*(3), 191-194

Leckey, M., & Federmeier, K. D. (2019). The P3b and P600 (s): Positive contributions to language comprehension. *Psychophysiology*, e13351. doi: 10.1111/psyp.13351

Lim, S.-J., Shinn-Cunningham, B. G., & Perrachione, T. K. (2019). Effects of talker continuity and speech rate on auditory working memory. *Attention, Perception, & Psychophysics, 81*(4), 1167-1177. doi: 10.3758/s13414-019-01684-w

Lim, S.-J., Tin, J. A., Qu, A., & Perrachione, T. K. (2019). *Attentional reorientation explains processing costs associated with talker variability.* Paper presented at the 19th International Congress of Phonetic Sciences. retrieved from http://sites.bu.edu/cnrlab/files/2019/05/Lim-Tin-Qu-Perrachione-2019-ICPhS-TCTA.pdf

Magnuson, J. S., & Nusbaum, H. C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology: Human Perception and Performance, 33*(2), 391. doi: 10.1037/0096-1523.33.2.391

Munro, M. J. (2003). A Primer on Accent Discrimination in the Canadian Context. *TESL Canada journal, 20*(2), 38-51

Münte, T. F., Heinze, H.-J., Matzke, M., Wieringa, B. M., & Johannes, S. (1998). Brain potentials and syntactic violations revisited: No evidence for specificity of the syntactic positive shift. *Neuropsychologia, 36*(3), 217-226. doi: 10.1016/S0028-3932(97)00119-X

Novick, J. M., Hussey, E., Teubner-Rhodes, S., Harbison, J. I., & Bunting, M. F. (2014). Clearing the garden-path: Improving sentence processing through cognitive control training. *Language, Cognition and Neuroscience, 29*(2), 186-217. doi: 10.1080/01690965.2012.758297

Nusbaum, H. C., & Morin, T. M. (1992). Paying attention to differences among talkers. In Y. Tohkura, E. Vatikiotis-Bateson, Y. Sagisaka (Ed.), *Speech perception, production and linguistic structure* (pp. 113-134). Tokyo: Ohmsha, Ltd

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31*(6), 785-806. doi: 10.1016/0749-596X(92)90039-Z

Rabovsky, M., Hansen, S. S., & McClelland, J. L. (2018). I like coffee with cream and dog? Change in an implicit probabilistic representation captures meaning processing in the brain. *bioRxiv*, 138-149

Regel, S., Coulson, S., & Gunter, T. C. (2010). The communicative style of a speaker can affect language comprehension? ERP evidence from the comprehension of irony. *Brain research, 1311*, 121-135. doi: /10.1016/j.brainres.2009.10.077

Romero-Rivas, C., Martin, C. D., & Costa, A. (2015). Processing changes when listening to foreign-accented speech. *Front Hum Neurosci*, 9, 167. doi:10.3389/fnhum.2015.00167

Rünger, D., Schwager, S., & Frensch, P. A. (2010). Across-task conflict regulation: A replication failure. *Journal of Experimental Psychology: Human Perception and Performance, 36*(1), 136-146. doi: 10.1037/a0017172

Sabourin, L., Stowe, L. A., & De Haan, G. J. (2006). Transfer effects in learning a second language grammatical gender system. *Second Language Research, 22*(1), 1-29

Sassenhagen, J., & Fiebach, C. J. (2019). Finding the P3 in the P600: Decoding shared neural mechanisms of responses to syntactic violations and oddball targets. *NeuroImage*, *200*, 425-436. doi: 10.1016/j.neuroimage.2019.06.048

Sassenhagen, J., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2014). The P600-as-P3 hypothesis revisited: Single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned. *Brain and language, 137*, 29-39. doi: 10.1016/j.bandl.2014.07.010

Schacht, A., Sommer, W., Shmuilovich, O., Martienz, P. C., & Martin-Loeches, M. (2014). Differential task effects on N400 and P600 elicited by semantic and syntactic violations. *PLoS One, 9*(3), e91226. doi: 10.1371/journal.pone.0091226

Schweinberger, S. R., & Neumann, M. F. (2016). Repetition effects in human ERPs to faces. Cortex, 80, 141-153. doi: 10.1016/j.cortex.2015.11.001

Schweinberger, S. R., Pfütze, E.-M., & Sommer, W. (1995). Repetition priming and associative priming of face recognition: Evidence from event-related potentials. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21(3), 722. doi: 10.1037/0278-7393.21.3.722

Shinn-Cunningham, B.G. 2008. Object-based auditory and visual attention. Trends Cogn. Sci. 12, 182-186. doi: 10.1016/j.tics.2008.02.003

Simon, J. R. (1969). Reactions toward the source of stimulation. *Journal of experimental psychology*, *81*(1), 174. doi: 10.1037/h0027448

Sommer, W., Leuthold, H., & Matt, J. (1998). The expectancies that govern the P300 amplitude are mostly automatic and unconscious. *Behavioral and Brain Sciences*, *21*(1), 149-150. doi: 10.1017/S0140525X98210958

Sommer, W., Matt, J., & Leuthold, H. (1990). Consciousness of attention and expectancy as reflected in event-related potentials and reaction times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(5), 902. doi: 10.1037/0278-7393.16.5.902

Stürmer, B., Leuthold, H., Soetens, E., Schröter, H., & Sommer, W. (2002). Control over location-based response activation in the Simon task: behavioral and electrophysiological evidence. *Journal of Experimental Psychology: Human Perception and Performance, 28*(6), 1345. doi: 10.1037/0096-1523.28.6.1345

Thothathiri, M., Asaro, C. T., Hsu, N. S., & Novick, J. M. (2018). Who did what? A causal role for cognitive control in thematic role assignment during sentence comprehension. *Cognition, 178*, 162-177. doi: 10.1080/02687038.2017.1350630

van Berkum, J. J. (2004). Sentence comprehension in a wider discourse: Can we use ERPs to keep track of things. *The on-line study of sentence comprehension: Eyetracking, ERPs and beyond*, 229-270

van Berkum, J. J., van den Brink, D., Tesink, C. M., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience, 20*(4), 580-591

van de Meerendonk, N., Indefrey, P., Chwilla, D. J., & Kolk, H. H. (2011). Monitoring in language perception: Electrophysiological and hemodynamic responses to spelling violations. *Neuroimage, 54*(3), 2350-2363. doi: 10.1016/j.neuroimage.2010.10.022

van de Meerendonk, N., Kolk, H. H., Chwilla, D. J., & Vissers, C. T. W. (2009). Monitoring in language perception. *Language and Linguistics Compass, 3*(5), 1211-1224. doi: 10.1111/j.1749-818X.2009.00163.x van Herten, M., Kolk, H. H., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Brain Res Cogn Brain Res, 22*(2), 241-255. doi:10.1016/j.cogbrainres.2004.09.002

Vissers, C. T., Chwilla, D. J., & Kolk, H. H. (2006). Monitoring in language perception: The effect of misspellings of words in highly constrained sentences. *Brain Res, 1106*(1), 150-163. doi: 10.1016/j.brainres.2006.05.012

Vissers, C. T. W., Kolk, H. H., van de Meerendonk, N., & Chwilla, D. J. (2008). Monitoring in language perception: evidence from ERPs in a picture–sentence matching task. *Neuropsychologia, 46*(4), 967-982. doi: 10.1016/j.neuropsychologia.2007.11.027

Vuong, L. C., & Martin, R. C. (2014). Domain-specific executive control and the revision of misinterpretations in sentence comprehension. *Language, Cognition and Neuroscience, 29*(3), 312-325. doi: 10.1080/01690965.2013.836231

Wong, P. C., Nusbaum, H. C., & Small, S. L. (2004). Neural bases of talker normalization. *Journal of Cognitive Neuroscience, 16*(7), 1173-1184. doi: 10.1162/0898929041920522

Xu, J., Abdel Rahman, R., & Sommer, W. (2019). Perceived language competence modulates criteria for speech error processing: evidence from event-related potentials. *Language, Cognition and Neuroscience*, 1-14. doi: 10.1080/23273798.2018.1562558

Xu, J., Abdel Rahman, R., & Sommer, W. (2021). Sequential Adaptation Effects Reveal Proactive Control in Processing Spoken Sentences: Evidence from Event-related Potentials. *Brain and Language*, 214, 104904. doi: 10.1016/j.bandl.2020.104904

Original Research Articles

The dissertation is based on the following original research articles.

Article I:

Xu, J., Abdel Rahman, R., & Sommer, W. (2019). Perceived language competence modulates criteria for speech error processing: evidence from event-related potentials. *Language, Cognition and Neuroscience*, 1-14. doi: 10.1080/23273798.2018.1562558

Article II:

Xu, J., Abdel Rahman, R., & Sommer, W. (2021). Sequential adaptation effects reveal proactive control in processing spoken sentences: Evidence from event-related potentials. Brain and Language, 214, 104904. doi: 10.1016/j.bandl.2020.104904

Article III:

Xu, J., Abdel Rahman, R., & Sommer, W. (2021). Who Speaks Next? Adaptations to Speaker Identity in Processing Spoken Sentences. *Manuskript unter Begutachtung für Psychophysiology*

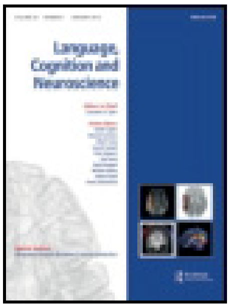
Original Research Articles

I. Perceived Language Competence Modulates Criteria for Speech Error Processing: Evidence from Event-related Potentials

Jue Xu, Rasha Abdel Rahman, & Werner Sommer

Reference:

Xu, J., Abdel Rahman, R., & Sommer, W. (2019). Perceived language competence modulates criteria for speech error processing: evidence from event-related potentials. *Language, Cognition and Neuroscience*, 1-14. doi: 10.1080/23273798.2018.1562558



Language, Cognition and Neuroscience

ISSN: 2327-3798 (Print) 2327-3801 (Online) Journal homepage: https://www.tandfonline.com/loi/plcp21

Perceived language competence modulates criteria for speech error processing: evidence from event-related potentials

Jue Xu, Rasha Abdel Rahman & Werner Sommer

LANGUAGE, COGNITION AND NEUROSCIENCE
https://doi.org/10.1080/23273798.2018.1562558

REGULAR ARTICLE

OPEN ACCESS

Perceived language competence modulates criteria for speech error processing: evidence from event-related potentials

Jue Xu, Rasha Abdel Rahman and Werner Sommer

Institut für Psychologie, Humboldt-Universität zu Berlin, Berlin, Germany

ABSTRACT

With event-related potentials we examined how speaker identity affects the processing of speech errors. In two experiments with probe verification and sentence correctness judgement tasks, respectively, grammatical agreement violations and slips of the tongue were embedded in German sentences spoken in native or Chinese accent. Portraits of European or Asian persons served as cues for speaker's identity. In Experiment 1, only a P600 was elicited by grammatical agreement errors in native speech in the second presentations. In Experiment 2, grammatical errors again elicited a P600 only in native speech. Slips of the tongue, however, elicited a P600 in both native and non-native speech and a N400 for native speech. Hence, perceived speaker nativeness seems to modulate the integration of grammatical agreement violations into the utterance. Slips of the tongue induced (re)interpretation processes (P600) for both native and non-native speech, whereas retrieval of lexico-semantic information (N400) is reduced in non-native speech.

ARTICLE HISTORY
Received 29 January 2018
Accepted 6 December 2018

KEYWORDS
Grammatical agreement violation; N400; P600; slips of the tongue; speaker identity

Introduction

Natural speech includes occasional errors, not only in second-language (L2) users but also in highly competent native speakers (L1 users). The present study aims to provide evidence from event-related potentials (ERPs) that such differences in perceived speaker competence may modulate criteria for processing speech errors. As criterion modulation may depend on the type of error, we separately considered grammatical agreement violations and slips of the tongue (mostly semantic blends).

Prior work has shown that speech perception actively uses context information about a speaker's identity to anticipate upcoming speech. For example, stereotype-driven inferences about sex, age or social status based on the talker's voice may trigger distinct brain responses when perceiving incongruent versus congruent speech input (Lattner & Friederici, 2003; van Berkum, van den Brink, Tesink, Kos, & Hagoort, 2008).

Differences in accent and frequent errors typically distinguish L2 speech from L1 speech. Non-native accent differs in segmental inventory (Munro, 2003) and prosodic aspects (Anderson-Hsieh, Johnson, & Koehler, 1992) from native phonological norms. Speech errors and especially grammatical errors are more frequent in L2 than L1 speech. Foreign language learners often have difficulties with gender agreement, especially when their L1 lacks grammatical gender (Franceschina, 2005; Sabourin, Stowe, & De Haan, 2006), for example learners of German whose L1 is Chinese, because the Chinese language does not have grammatical morphology for marking number, gender and case (Chen, Shu, Liu, Zhao, & Li, 2007). Chinese speakers of German are therefore more likely to produce grammatical agreement violations than native speakers of German. During face-to-face communication, when expecting non-native speech, listeners have to take into account such errors and the foreign accent. This expectation should modulate processing criteria for syntax errors in non-native versus native speech.

Slips of the tongue, like Spoonerisms, such as "Our queer old dean" rather than "Our dear old queen", are frequently encountered every-day speech errors. In German there are five major types of slips of the tongue: blends, exchanges, anticipations, postpositions, and substitutions, which could affect language units of different sizes, from syllables, words, phrases, up to whole syntactic structures (Meringer & Mayer, 1895). Despite being of great interest for the study of speech production and comprehension, the neural correlates of perceiving slips of the tongue and their relationship with native or non-native speaker identities, are not yet fully understood.

CONTACT Jue Xu xujue@hu-berlin.de
Supplemental data for this article can be accessed https://doi.org/10.1080/23273798.2018.1562558.
This article has been republished with minor changes. These changes do not impact the academic content of the article.

The EEG is widely used to examine language comprehension. Prior work identified two ERP components correlated with processing semantic and syntactic information of speech: the N400 and the P600 component. The N400 component is a negative voltage deflection peaking around 400 ms at centro-parietal sites, is taken to reflect semantic processing and context integration of verbal and non-verbal stimuli (Kutas & Federmeier, 2011; van Berkum, 2004). This component has also been taken to reflect prediction error (Rabovsky, Hansen, & McClelland, 2018). The P600 is a positive component maximal at centro-parietal sites starting around 500 ms, typically extending to 800 ms or more, which was initially associated with syntactic processing, but was later observed also in response to thematic and other semantic violations, without necessarily eliciting a preceding N400 effect (see Kuperberg, 2007, for a review). In their Retrieval-Integration (RI) account of language processing, Brouwer, Crocker, Venhuizen, and Hoeks (2017) recently suggested that the N400 amplitude reflects activation and retrieval of lexico-semantic information from long-term memory and the P600 component indicates the integration of the activated information into online utterance interpretation.

The majority of earlier ERP studies on accented speech processing focused on how lexico-semantic violations or grammatical errors are perceived differently in native, foreign and regional accents. Based on knowledge about frequent or infrequent error types as a function of speaker identity, neural correlates of syntactic processing may change (e.g. Grey & van Hell, 2017; Hanulíková, van Alphen, van Goch, & Weber, 2012; Romero-Rivas, Martin, & Costa, 2015). For example, Hanulíková et al. (2012) tested gender agreement violations and semantic world knowledge violations in native and Turkish-accented Dutch. They found a P600 effect to gender errors in L1 speech but not in L2 speech, whereas comparable N400 effects were elicited by semantic anomalies in L1 and L2 speech.

Romero-Rivas et al. (2015) also explored how semantic world knowledge violations were processed in Spanish spoken in native speech and with four different foreign accents (French, Greek, Italian, Japanese). An N400 effect was elicited by semantic violations in native speech followed by a late positivity, while only an N400 effect was found in non-native speech. They suggested that listeners avoid trying to find an alternative meaning for the semantic violations in non-native speech; hence, no re-analysis was carried out.

The current study intended to provide further evidence on how native or non-native speaker identities affect the processing of grammatical errors, and to explore the neural correlates of perceiving slips of the tongue in continuous speech and whether these correlates would be modulated by speaker identity.

Outline of experiments and predictions

Faces as cues

In order to allow listeners to derive predictions before language processing, we used faces as visual cues providing explicit advance information whether native or non-native speech would be presented. It is natural in daily communication that interlocutors retrieve information about each other from appearance before the conversation. The studies mentioned above presented auditory sentences without any previous cues about speaker identity; hence, only after listeners recognised the non-native accent as an indexical property of the speaker, could processing of incoming signals begin to differ. However, individuals differ in their ability to recognise different accents. This could lead to different ERPs in response to the errors. Indeed, Grey and van Hell (2017) found an N400-like effect to English subject pronoun errors only in a subset of listeners that correctly identified the foreign accent. We relied not only on previous visual cues but also on accents, in which native and non-native accent was associated with native and non-native facial appearance, respectively.

Speech errors

Grammatical agreement violations. The current study focused on grammatical agreement violations with the following syntactic patterns (Appendix C): gender agreement violation between determiner and noun (e.g. Barber & Carreiras, 2005; Hagoort, 2003; Molinaro, Vespignani, & Job, 2008), number agreement violation between subject/pronoun and verb (e.g. Hagoort & Brown, 2000; Roehm, Bornkessel, Haider, & Schlesewsky, 2005) or between determiner and noun (e.g. Hagoort, 2003), and case agreement violation between verb and object (e.g. Roehm et al., 2005). All violations should elicit a P600 effect when spoken by native speakers (see Molinaro, Barber, & Carreiras, 2011, for a review). As shown by Romero-Rivas et al. (2015) and Hanulíková et al. (2012), listeners seem not to try re-interpreting syntactic and semantic errors made by L2 speakers. Thus, we predicted that grammatical agreement violations in non-native speech would not engender a P600 effect.

Slips of the tongue. In the current study, blends were used to represent slips of the tongue. Blends are generated because of the similarity in meaning or form of the derived sentences, phrases or words (Meringer & Mayer, 1895). The root words or phrases of blends used in the current study share semantic meaning under the same context.

All blends used in our sentences differed from the intended correct versions only in one content word. Superficially, they were either pseudo-words constructed by recognisable word fragments or illegal constitutes in phrasal structures. The blends in the materials were realised on two levels (see examples in Table 1). Either two different words (root words) were blended into one word (blend on word-level) as in Example (i), in which *aufgeschwächt* is blended from *aufgeweicht* [softened] and *geschwächt* [weakened], or two phrases (root phrases) were blended into one phrase (blend on phrase-level), as in Example (ii), in which *j-m ein Schnippchen spielen* is blended from *j-m ein Schnippchen schlagen* [cheat someone] and *j-m einen Streich spielen* [play a trick on someone]. The resultant blends were illegal in the whole sentence frame either because they were pseudo-words like *aufgeschwächt*, or because they created illegal phrase structures as shown in Example (ii).

In contrast to the well-investigated effects of grammatical agreement violations on the P600 component, the situation is less clear for slips of the tongue. We hypothesised that a P600 effect would only be engendered by such errors in native speech and an N400 effect would only be engendered by such errors in non-native speech, explained separately for the two types below.

Critically, word-level blends and their correct versions shared the same initial phoneme(s). ERPs were time-locked to the divergence points of these two conditions, where the blending word and the corresponding correct word started to acoustically diverge from each other, as defined by van Petten, Coulson, Rubin, Plante, and Parks (1999) and van den Brink, Brown, and Hagoort (2001). Both studies and Connolly and Phillips (1994) reported a delayed latency of the N400 effect in semantically anomalous conditions with the same initial phonemes as the congruent words with ERPs time-locked to word onset. Therefore, the N400 component is related to the moment, at which the acoustic input first diverged from expectation.

Table 1. Examples of blends.

	(i)
a.	Der Bund ist von der$_{[f]}$ Reform$_{[f]}$ stark aufgeweicht worden.
	(The bund has been greatly weakened by the$_{[f]}$ reform$_{[f]}$.)
b.	Der Bund ist von dem$_{[m]/[n]}$ Reform$_{[f]}$ stark aufgeweicht worden.
c.	Der Bund ist von der Reform stark aufgeschwächt worden.
	(ii)
a.	Er hat ihnen ein$_{[n]}$ Schnippchen$_{[n]}$ geschlagen mit der Erbschaft.
	(He played a$_{[n]}$ trick$_{[n]}$ on them with the inheritance.)
b.	Er hat ihnen einen$_{[m]}$ Schnippchen$_{[n]}$ geschlagen mit der Erbschaft.
c.	Er hat ihnen ein Schnippchen gespielt mit der Erbschaft.

Notes: In each example, a. is well-formed, b. contains a grammatical agreement violation, and c. contains a blend. English translations of a. are given in the same font style in brackets. Single- and wavy-underlined words are triggers for grammatical agreement violations and blends, respectively. Grammatical gender (m = masculine, f = feminine, n = neuter) refers to the gender of this noun if subscripted under a noun; otherwise, it refers to the correct gender that the determiner should lead.

As suggested by Pickering and Garrod (2013), language comprehension anticipates upcoming words at different linguistic levels. Based on context information and the early processing of initial sounds of the word, multiple lexical candidates would be activated online, where both word form and context information contribute to the retrieval of semantic information (van den Brink et al., 2001). For a word-level blend, the acoustic-phonological processing of the initial acoustic input and the lexical selection of multiple candidates should be successful. Since the remaining word fragments of the blends are indeed parts of other suitable candidates, their word form information would also be activated. Therefore, no further retrieval of lexico-semantic information should be needed for word-level blends in native speech, not yielding any N400 effect.

Phrase-level blends were realised by substituting one word in a phrase by a word from another phrase. Although failing to build a correct syntactic hierarchy, the substitute should not be considered as semantic anomaly, because it carries suitable semantic information from the two root phrases. No further semantic information needs to be retrieved; hence, no N400 was expected.

Both kinds of blends in native speech used here should elicit a P600 effect, reflecting a mechanism of repair and integration of activated information into online utterance interpretation, as suggested in the RI theory (Brouwer et al., 2017). In line with this idea, van Herten, Kolk, and Chwilla (2005) found only a P600 effect but no N400 in response to semantic reversal anomalies like "The *cat* that fled from the *mice* ran across the room" (translation of the original Dutch sentence). They interpreted the P600 as a monitoring component that checks upon the veridicality of one's sentence perception. In conclusion, for slips of the tongue in native speech, we predicted a P600 effect but no N400.

Another key issue concerned whether there would be a difference in the perception of slips of the tongue between native and non-native speech. It is not clear, whether slips of the tongue are indeed more expected in native than non-native speech. We hoped to provide some evidence in this regard too. Regarding the re-interpretation process, our hypotheses for blends in non-native speech were similar to grammatical errors: no P600 effect, reflecting reduced or no effort in repairing errors made by L2 speakers.

We expected an N400 effect engendered by blends in non-native but not in native speech. The main reason for this difference was the foreign accent. As suggested by Pickering and Garrod (2013), the comprehension system may use the production system to covertly imitate the speaker and anticipate upcoming speech in communication. The increased phonetic variability and lower reliability in foreign-accented speech may cause unsuccessful or reduced lexical activation. Therefore, we hypothesised that increased lexico-semantic retrieval would be needed for blends in non-native speech, reflected in an N400 effect.

In a nutshell, the hypothesis of the current study was that listeners interpret errors partially depending on who is speaking. In particular, we expected a P600 effect to blends in native speech, and an N400 effect to blends in non-native speech. Grammatical agreement violations were expected to engender a P600 effect in native but no effect in non-native speech.

Further questions

As a further question we asked whether short-term experience with speech errors and accents would modulate their processing. We introduced a second experimental block repeating the sentences of a first block in a different order. Hanulíková et al. (2012) split the data into the first and second halves of their experiment and found a P600 effect to native grammatical errors only in the first half. Experience with a given speaker identity, in their case the constant number of errors in both speaker identities, might affect the stereotype about the speaker. We expected to find an attenuated P600 to native errors in Block 2 compared to Block 1. In addition, Romero-Rivas et al. (2015) showed that listeners improved at recognising, retrieving and integrating incoming words after brief exposure to foreign-accented speech. Listeners can quickly adapt to foreign-accented speech and the comprehension generally improves over time (Cristia et al., 2012). We therefore expected an emerging P600 effect in non-native accented speech in Block 2 compared to Block 1.

Considering that listeners may be amused by speech errors, we also applied electromyographic (EMG) electrodes over the *M. zygomaticus major* (Fridlund & Cacioppo, 1986) to detect dynamic smiles during the test, possibly elicited by the speech errors.

Experiment 1

Methods

Participants

A total of 27 participants were tested. Two of them were excluded from analysis because of excessive error rates in the probe verification task (22.2% and 30.6%), and one because of ambidexterity (final sample: 16 women and 8 men, mean age = 26 years, range: 18–36). All participants were native German speakers without hearing, neurological, or psychiatric disorders and with normal or corrected-to-normal visual acuity and normal colour vision according to self-report. They were right-handed according to the Edinburgh Questionnaire (Oldfield, 1971), gave informed consent and received payment or course credits for participation. None of the participants was of Asian ethnic background or reported knowledge of an Asian language. All tests were carried out at the psychology department in Humboldt-Universität zu Berlin.

Materials

A total of 180 German sentences were constructed (mean length = 7.78 words, $SD = 1.89$), containing slips of the tongue, taken from Leuninger (1996, 1999) and the online blog of Wietzel-Winkler (2017). All slips of the tongue were content words (nouns: 49.44%, verbs: 31.67%, adjectives/adverbs: 18.89%). In Experiment 1, we also presented phonological slips of the tongue (20%) together with the blends (80%), for example, “Die *Piratendatei* wurde 2006 in Berlin gegründet” [The *Pirate File* was founded in 2006 in Berlin], where the intended word “Piratenpartei” [Pirate party] was mispronounced as *Piratendatei [Pirate file]* because the activated syllable “de” in “wurde” [was] was inserted into the intended word plan.

The two kinds of speech errors in our materials did not overlap with each other. Grammatical agreement violations affected either a verb or a noun in the size of inflectional morphemes, while the blends were distinguished from the intended words at the size of several syllables up to a word. Sentences with blends accorded all correctly to grammatical agreements in German. A full list of stimuli can be found in Appendix A.

We collected information on word length (letter and syllable number) and word frequency (based on lemma) of all critical words from the online German linguistic corpus dlexDB (Heister et al., 2011). One-factor ANOVAs with factor letter number, syllable number and word frequency were carried out separately to compare the two root conditions. No significant differences were found ($Fs \leq 3.41$, $ps \geq .066$).

From each well-formed critical word, that corresponded to a slip of the tongue, one further version was derived that contained a grammatical agreement violation in gender (63.33%), number (28.33%), or case (8.33%), resulting in 180 sentence triplets with critical words that were well-formed, contained a slip of the tongue or grammatical agreement violation. No critical word in any sentence was at the first or last word position.

All 540 sentences were spoken by two female speakers, a native German speaker pronouncing in standard

German and a native Chinese speaker speaking Chinese-accented German, with neutral intonations at normal speed. A total of 1080 audio files were recorded in a professional studio using a Neumann® TLM 103 condenser microphone with fixed heart-shaped directivity. Sentences were digitised with 44.1 kHz at 24 bit resolution and stored in wave-format. GoldWave® v5.70 software was used to change the pitch of both speakers into 15 different voices and to mark the onsets of critical events in each sound file. Each sentence pair spoken by the two speakers was normalised according to their mean duration. Mean sentence duration was 3.2 s (*SD* = 0.73) and did not vary across the native and non-native speaker conditions.

For grammatical agreement violations and their corresponding correct versions, markers for later EEG segmentation were placed at the onsets of critical words where the ungrammaticality became apparent. For slips of the tongue, 111 out of the 180 sentences (61.67%) had a critical word that shared the same first syllable(s) with its corresponding correct version. As explained above, ERPs were time-locked to their divergence points.

Design

The experiment used a 2 × 2 design: native or non-native speaker identity and 2 error types – grammatical agreement violations and slips of the tongue. The 1080 audio files were divided into 6 subsets. Only one version of each sentence triplet appeared in one subset. Half of the sentences in a given subset were non-native accented and half were native accented. Within a test session, one subset of 180 audio files was presented twice in two separate blocks with different randomised orders. All sentences and conditions were thus fully counterbalanced across each subgroup of six participants.

Pictures of 90 Caucasian and 90 Chinese female faces represented 180 speaker identities from two different ethnic backgrounds. European faces were taken mostly (*N* = 85) from the FACES database (Ebner, Riediger, & Lindenberger, 2010; Lindenberger, Ebner, & Riediger, 2005–2007), and the others from the Radboud Faces Database (Langner et al., 2010). Chinese face pictures were taken from the CAS-PEAL face database (Gao et al., 2008). All faces showed neutral expressions with direct gaze at the viewer. All pictures were converted in Adobe Creative Suite 6® Photoshop into grey scale and cut into square format with only the face filling the square. Each face was assigned to two sentence triplets. The assignment of face to voice was fixed and did not change across the experiment.

Apparatus

The computer monitor used in the test was 19-inch DELL® 1908 FPb. The audio files were presented using two Creative® Gigaworks T20 loud speakers placed at both sides of the monitor.

Procedure

Participants were tested in a sound-attenuated chamber. Audio volume was adjusted to a clear and comfortable level for each participant before the experiment. Each trial began with a fixation cross presented in the middle of the screen for 1 s, followed by a face picture. After 800 ms, the audio signal started, while the picture remained on the screen. One second after the end of the sentence, a blank screen was presented for 200 ms. There were breaks every 45 trials of participant-determined duration.

In 10% of all trials (*N* = 36), randomly interspersed and equally distributed across blocks, a probe verification task was included. After the presentation of the face, a noun appeared on the screen. Half of these nouns referred to concepts in the preceding sentence. For example, for the sentence "Mutti sagt, dass die Milch bei Gewitter schnell sauer wird" [Mom says that milk will deteriorate quickly during thunderstorms], the probe word was "Wetter" [Weather]. Participants had to decide whether or not the noun had been referred to in the sentence content by pressing one of two buttons placed on the table in front.

Participants were instructed to avoid movements during the experiment and not to blink while the face was shown. They were instructed to fixate the visual stimuli, pay attention to the pictures and listen to the sentences for understanding. Accents and speech errors were not mentioned in the instructions. After the experiment, a short calibration procedure obtained prototypical eye movements artefacts, to be later used for correction. Finally, participants filled in a questionnaire about the intelligibility of the sentences and the foreign accent (Appendix B).

Electrophysiological recordings

The continuous EEG was recorded from 64 Ag/AgCl electrodes arranged according to the extended 10/20 system. The left mastoid was used as initial reference. We used electrodes near the left and right canthi of both eyes and above and beneath the left eye to register eye movements and blinks. In addition, two Ag/AgCl electrodes, 4 mm in diameter, were positioned over the zygomaticus major on the right side of the face in order to detect smiles or laughter in response to errors. Impedances of all electrodes were kept below 5 kΩ.

The raw EEG and EMG signals were amplified and filtered online at a band pass of 0.1–1000 Hz at an initial sampling rate of 5000 Hz converted to 500 Hz by BrainAmp ExG amplifier (Brain Products®). Offline, the EMG was rectified and filtered with 30 Hz high-pass (12 dB/oct) and a moving-average filter integrating over 30 ms. The EEG was re-calculated offline to average reference and low-pass filtered at 30 Hz (24 dB/oct). Eye movement and blink artefacts were corrected employing BESA® software (Berg & Scherg, 1994). The EEG and EMG data were segmented into epochs of 1.3 s, starting 100 ms before the onset of the critical events; these 100 ms were used as baseline. EEG segments with a voltage range exceeding 100 µV were excluded using automatic artefact rejection. Finally, segments were averaged separately for each condition, block, electrode, and participant. All EEG processing steps were conducted using the MATLAB® R2016a software and the toolboxes EEGLAB (Delorme & Makeig, 2004) and FieldTrip (Oostenveld, Fries, Maris, & Schoffelen, 2011), and all EMG processing was conducted with BrainVision Analyzer 2.1 (Brain Products®) in a 64-Bit Windows® 7 operating system.

Data analysis

Mean amplitudes of the EMG segments between 300 and 600 ms were calculated for each participant and entered into an ANOVA with repeated measures on factors error type (slip of the tongue, grammatical agreement violation), well-formedness (erroneous, well-formed), and speaker identity (native, non-native).

Mean ERP amplitudes in a centro-parietal ROI of 25 electrodes (C1/2, C3/4, CP1/2, CP3/4, CP5/6, P3/P4, P5/6, PO3/4, PO7/8, O1/2, Cz, CPz, Pz, POz, Oz) in two time windows were analysed with repeated measures ANOVAs. Informed by previous research about speech perception in the auditory modality (e.g. Hanulíková et al., 2012; Koester, Gunter, Wagner, & Friederici, 2004; Mueller, Oberecker, & Friederici, 2009; Romero-Rivas et al., 2015, 2016; Rossi, Gugler, Hahne, & Friederici, 2005), we established a time window of 600–1200 ms for the P600 component. For the N400 component, we selected a time window of 300–500 ms based on previous literature (e.g. Kutas & Federmeier, 2011). For grammatical agreement violations, one three-way ANOVA with speaker identity (native, non-native), well-formedness (erroneous, well-formed), and block (Block 1 and Block 2) as within-subjects factors was done in P600 time window. For slips of the tongue, two three-way ANOVAs with the same factors were conducted separately for the N400 and P600 time windows. In addition, we used the Bonferroni correction for *post hoc* analyses.[1]

Results

Behavioural results

According to the post-experimental questionnaires, all participants reported to have understood at least 90% of the sentences. Twenty-two participants identified the foreign accent as Chinese or Asian, and two participants had no idea about its regional origin.

Mean error rate in the probe verification task was 9.49% (mean error number = 3.5, $SD = 1.7$). To check whether the error rate was affected by the accent or error type, an ANOVA with repeated measures including factors speaker identity (native, non-native) and sentence type (slips of the tongue, grammatical agreement violations, well-formed versions) was conducted. No significant effect or interaction was found ($Fs < 1$).

Electrophysiological results

EMG results. ANOVA on the *zygomaticus* data did not reveal any significant main effect or interaction ($Fs \leq 1.67$, $ps \geq .209$).

EEG results. The three-way ANOVA regarding the grammatical agreement violations revealed a three-way interaction of factors block, speaker identity and well-formedness ($F(1, 23) = 4.64$, $p = .042$, $\eta_p^2 = .168$). Follow-up pairwise comparisons revealed a significant P600 effect for native speakers in Block 2 ($F(1, 23) = 5.71$, $p = .025$, $\eta_p^2 = .199$). No other effects were found ($Fs \leq 2.83$, $ps \geq .106$).

For slips of the tongue, the ANOVA in the N400 window revealed a marginally significant effect of well-formedness ($F(1, 23) = 4.05$, $p = .056$, $\eta_p^2 = .150$) and its interaction with block ($F(1, 23) = 2.97$, $p = .098$, $\eta_p^2 = .114$). As can be seen in Figure 1, the ERP difference waveforms indicate that slips of the tongue in native speech elicited a negativity around 300–500 ms relative to well-formed versions, possibly an N400 effect, which was absent in the difference waveforms in non-native speech. Therefore we performed a *post hoc* pairwise comparison between speaker identity and well-formedness on this effect. This analysis confirmed that the effect was significant in native speech ($F(1, 23) = 4.55$, $p = .044$, $\eta_p^2 = .165$) but not in non-native speech ($F(1, 23) = .24$, $p = .632$, $\eta_p^2 = .010$).

In the ANOVA regarding the P600 effect for slips of the tongue, the factor speaker identity was significant ($F(1, 23) = 4.68$, $p = .041$, $\eta_p^2 = .169$). No other effects or interactions were found ($Fs \leq .01$, $ps \geq .118$), even though the P600 component was larger in the erroneous than in the well-formed conditions (see ERP difference waves in Figure 1).

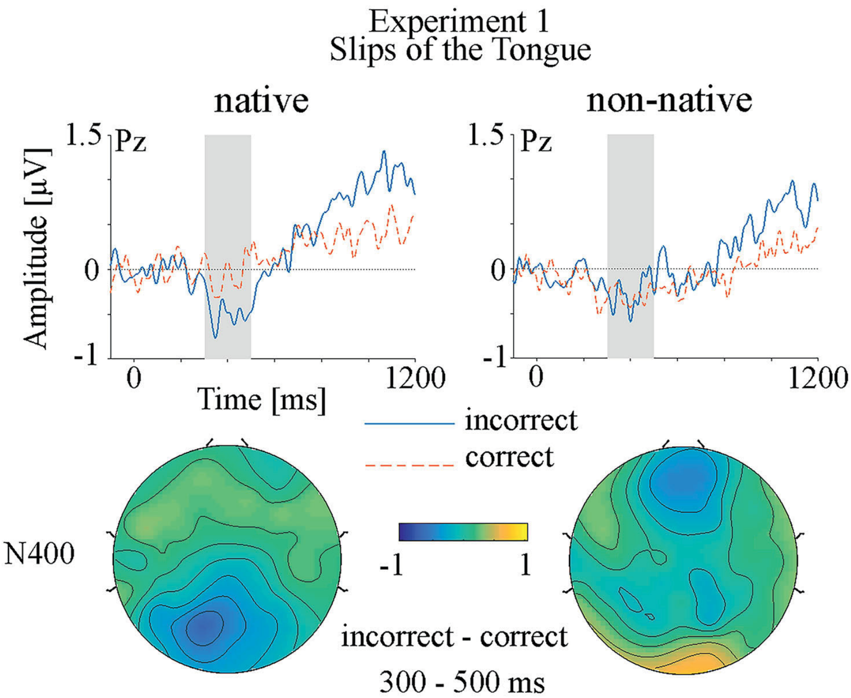


Figure 1. N400 Effect triggered by Slips of the Tongue in Experiment 1. Note: Grand-average difference topographies represent difference maps of erroneous minus well-formed versions separately averaged for native and non-native speaker conditions in 300–500 ms time window. ERPs represent grand means ($N = 24$) at electrode Pz separately averaged for native and non-native speaker identity conditions. Positive is plotted upward. Time window for the N400 effect is shaded.

Discussion

Grammatical errors evoked a P600 effect only in native speech and only in Block 2. It was in line with our expectation that grammatical errors would only engender a P600 effect in native but not in non-native speech. However, the result that this effect in native speech was absent in Block 1 and emerged in Block 2 was different from Hanulíková et al. (2012), who found the P600 effect to be present only in the first half of their experiment. Normally, when sentences are repeated, it should be easier and less effortful to process them. However, the P600 effect increased in the second presentation. Possibly, a reinterpretation of the sentences with errors was enhanced after the listeners had accumulated enough experience with this type of mistakes. The repetition in Block 2 could also have primed certain errors. This issue is further elaborated in the Discussion of Experiment 2.

Even though the averaged ERP amplitudes and topographies indicated a P600 effect elicited by slips of the tongue in both speaker identities, this was not statistically confirmed. The P600 effect to both kinds of errors seemed to have been greatly attenuated under this experimental design. It could be due to the task-sensitivity of the P600 component or to the high proportion of errors within the whole experiment (66%). As pointed out by Molinaro et al. (2011), the P600 amplitude is sensitive to the task and the proportion of violations in the whole experiment. Gunter and Friederici (1999) compared two types of syntactic errors in grammatical judgement task and physical judgement task. With the former task, verb inflection errors and word category errors both elicited robust N400 and P600 components, whereas with the latter task both components were greatly attenuated or absent for verb inflection errors and slightly diminished for word category violations. They suggested that the P600 reflects a relatively controlled language-related process. Hahne and Friederici (1999) found no P600 for phrase structure violations anymore after replacing a correctness judgement with a semantic coherence judgement task. Schacht, Sommer, Shmuilovich, Martinez, and Martin-Loeches (2014) repeated the Martín-Loeches, Nigbur, Casado, Hohlfeld, and Sommer (2006) study by replacing the original correctness judgement task by a probe verification task and found that the P600 disappeared while the N400 was only slightly smaller in amplitude under the indirect task.

Interestingly, we found a trend that slips of the tongue engendered an N400 effect. A *post hoc* comparison indicated the presence of an N400 effect in native but not in non-native speech. This effect seemed to be small and unstable across speaker identities. This could be due to a high variability of the materials that included 20% phonological slips of the tongue in addition to the 80% semantic blends.

In order to get a clearer view, we conducted Experiment 2, with three main changes relative to Experiment 1. First, we excluded phonological slips of the tongue and focused on blends to have a homogeneous set of stimuli. Second, instead of a probe verification task we used sentence correctness judgements for which the violations are directly task-relevant. We expected more pronounced P600 effect in Experiment 2, whereas little differences were expected for the N400 component, which seems to be more robust against task factors (Schacht et al., 2014). Third, to enhance the significance of errors for the listener, the overall proportion of errors in the speech material was decreased from 66% to 50%.

Experiment 2

Methods

Participants

A total of 26 new participants, selected according to the same criteria as in Experiment 1, were tested. Data of two persons had to be discarded because of either low judgement accuracies (79.0% for native and 53.8% for non-native speech) or high artefact rate in EEG data (21.63%) (final sample: 20 women and 4 men, mean age = 24 years, range: 18–42).

Materials

From the original 180 sentences with slips of the tongue, 135 sentences containing semantic blends were selected. In sentence versions with grammatical agreement violations, 63.70% were violations in gender, 23.70% in number, and 12.59% in case. Correct versions of the remaining 45 sentences were used as filler items. The same audio files were used as test materials (135 triplets × 2 speaker identities = 810 audio files as critical items; 45 correct sentences × 2 speaker identities = 90 audio files as fillers). Mean sentence duration of the critical items was 3.3 s (*SD* = 0.75) and did not vary across speaker conditions.

Design

Same as in Experiment 1, with the following changes. The 810 audio files were divided into 6 subsets: three subsets contained 88 native and 92 non-native sentences, and three subsets contained 88 non-native and 92 native sentences, and only one version of each triplet was present in one given subset. Each participant was presented with one subset and 45 correct fillers, which was either 22 native and 23 non-native, or reversed, to match the number of each accent in each subset, resulting in 50% error proportion for both speaker identities in every test. All sentences and conditions were thus fully counterbalanced across each subgroup of six participants.

Fifteen faces from each ethnic background were selected from the faces used in Experiment 1. A given face was consistently assigned to only one pitch (voice) throughout the experiment.

In the sentence correctness judgement, participants judged the overall correctness of the sentence directly after its presentation.

Procedure, apparatus and electrophysiological recordings

Same as in Experiment 1, except as follows. First, the fixation cross at the beginning of each trial was presented for 0.5 s. Second, participants were instructed to press one of two buttons within three seconds after the audio finished. Half of the participants pressed the left button for correct and the other button for incorrect sentences; for the other participants the assignment was reversed. After a button press or when three seconds had elapsed, the screen went black for 0.5 s, and the next trial began. Third, every 20 trials there was a break of participant-determined duration.

Data analysis

The accuracy of the correctness judgements, including both hits and correct rejections, were entered into an ANOVA with factors speaker identity (native, non-native) and sentence type (blends, grammatical agreement violations, and well-formed versions).

Raw EMG and EEG data were pre-processed and analysed in the same way as described for Experiment 1.

Results

Behavioural results

According to the post-experimental questionnaires, all participants correctly identified the foreign accent as either Chinese or Asian.

Mean accuracies of correctness judgements were 87.47% (*SD* = 3.94%) for native speech, and 78.52% (*SD* = 5.52%) for non-native speech (Figure 2). ANOVA revealed significant main effects of speaker identity ($F(1, 23) = 37.50$, $p < .001$, $\eta_p^2 = .620$), and error type (F(2,

46) = 28.78, $p < .001$, $\eta_p^2 = .556$), and an interaction of both factors ($F(2, 46) = 4.40$, $p = .018$, $\eta_p^2 = .160$). Follow-up analyses of the interaction showed no difference in the accuracies of judging blends in native and non-native speech ($F(1, 23) = 1.63$, $p = .215$), but accuracy was significantly higher for native compared to non-native speech containing grammatical agreement violations ($F(1, 23) = 17.83$, $p < .001$, $\eta_p^2 = .437$), or being well-formed ($F(1, 23) = 24.07$, $p < .001$, $\eta_p^2 = .517$).

Electrophysiological results

EMG results. ANOVA on the *zygomaticus* data did not reveal any significant main effect or interaction ($Fs \leq 2.15$, $ps \geq .131$).

EEG results. In the three-way ANOVA for grammatical agreement violations, there were significant main effects of block ($F(1, 23) = 7.96$, $p = .010$, $\eta_p^2 = .257$), well-formedness ($F(1, 23) = 7.90$, $p = .010$, $\eta_p^2 = .256$), and speaker identity ($F(1, 23) = 5.98$, $p = .023$, $\eta_p^2 = .206$). Well-formedness interacted with block ($F(1, 23) = 4.50$, $p = .045$, $\eta_p^2 = .164$) and with speaker identity ($F(1, 23) = 10.26$, $p = .004$, $\eta_p^2 = .308$). Follow-up analyses on the interaction between well-formedness and speaker identity revealed a significant effect of well-formedness for native speakers ($F(1, 23) = 17.14$, $p < .001$, $\eta_p^2 = .427$) but none for non-native speakers ($F(1, 23) = .732$, $p = .401$). Follow-up analyses on the interaction between well-formedness and block revealed a significant effect of well-formedness in Block 2 ($F(1, 23) = 13.67$, $p = .001$, $\eta_p^2 = .373$) but not in Block 1 ($F(1, 23) = 1.47$, $p = .237$). Visual inspection of the topographies and the difference waves confirmed that there was a P600 effect elicited by grammatical agreement violations in native speech, which was absent in non-native speech (Figure 3).

For slips of the tongue, the ANOVA of N400 effects revealed a significant effect of block ($F(1, 23) = 6.57$, $p = .017$, $\eta_p^2 = .222$) and a significant interaction between well-formedness and speaker identity ($F(1, 23) = 5.23$, $p = .032$, $\eta_p^2 = .185$). Follow-up analyses on this interaction confirmed that well-formedness was only significant in native speech ($F(1, 23) = 5.16$, $p = .033$, $\eta_p^2 = .183$) but not in non-native speech ($F(1, 23) = 1.10$, $p = .306$, $\eta_p^2 = .046$). As can be seen in Figure 4, slips of the tongue in native-speech resulted in a larger N400 compared with correct sentences, which was absent in non-native speech.

The three-way ANOVA of the P600 effects for slips of the tongue revealed a significant effect of block ($F(1, 23) = 8.43$, $p = .008$, $\eta_p^2 = .268$) and well-formedness ($F(1, 23) = 22.21$, $p < .001$, $\eta_p^2 = .491$). There was also a marginally significant interaction between speaker identity and well-formedness ($F(1, 23) = 2.98$, $p = .098$, $\eta_p^2 = .115$).

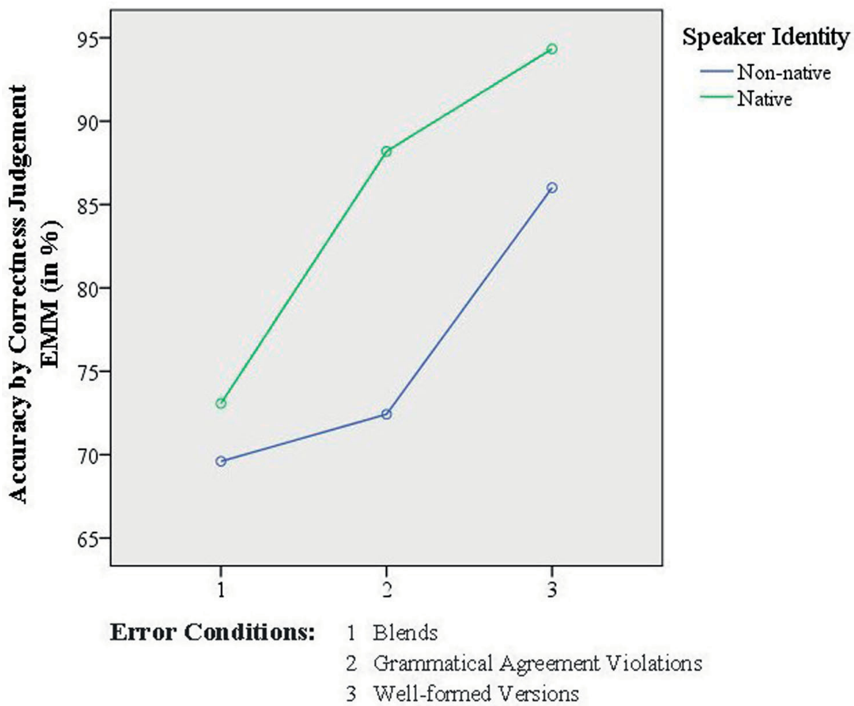


Figure 2. Accuracy by Correctness Judgement Task.

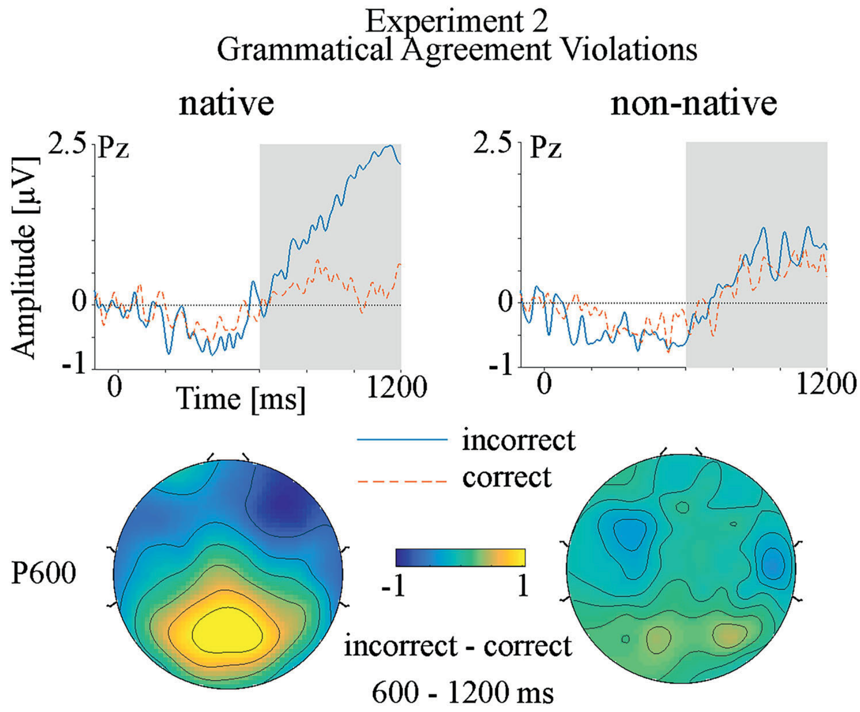


Figure 3. P600 Effect triggered by Grammatical Agreement Violations in Experiment 2. (Contrast and formations in the same way as Figure 1.)

Follow-up analyses on this interaction showed that well-formedness was significant in both native speaker identity ($F(1, 23) = 15.43$, $p = .001$, $\eta_p^2 = .401$) as well as non-native identity ($F(1, 23) = 4.93$, $p = .037$, $\eta_p^2 = .176$). Visual inspection of the topographies and the difference waves confirmed a P600 effect elicited by slips of the tongue in both native and non-native speech (see also Figure 4).

Discussion

Grammatical agreement violations

In Experiment 2 with a sentence correctness judgement task, grammatical agreement violations elicited a P600 effect that was only present in native speech perception, which is in line with the results in Hanulíková et al. (2012) and Romero-Rivas et al. (2015), indicating that listeners re-interpret these errors only for native speech.

In spoken language perception, word form information is mostly conveyed phonologically. A non-native accent made it more difficult for listeners to recognise words in a bottom-up way. What's more, stereotypical beliefs would suggest that L2 speakers have difficulties meeting grammatical agreements in natural speech. Hence, such errors are more expected from non-native speakers. Grammatical agreement errors are actually errors in word forms realised in inflectional morphemes, which don't necessarily hinder retrieving and apprehending the core meaning of the utterance. The non-native accent and the expectation of word form errors may have rendered the L2 speech seem less suitable for a bottom-up strategy based on word form information. Hence, for the sake of a more efficient communication with non-native speakers, listeners may have adapted a strategy that actively suppressed processing word forms and concentrated on interpreting the approximate meaning of the utterance and intention of the speaker.

Slips of the tongue

In Experiment 2, slips of the tongue elicited a P600 effect in both native and non-native speech, while an N400 effect was present for such errors only in native speech. In Romero-Rivas et al. (2015), both effects were elicited by semantic violations in native speech, while only an N400 effect but no P600 existed in non-native speech. Our results indicate that blends in native speech are processed in a similar way as semantic violations (with an N400 and a P600 effect), but blends in non-native speech are processed differently from pure semantic violations, eliciting only a P600 effect.

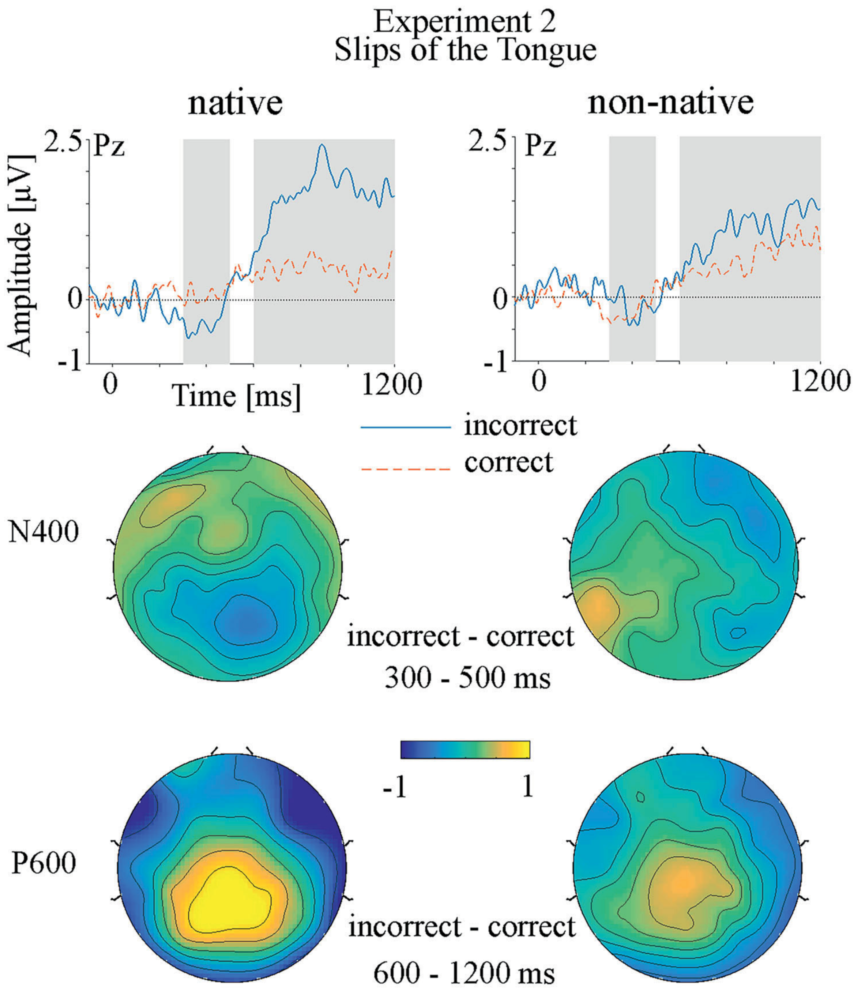


Figure 4. N400 and P600 Effects triggered by Slips of the Tongue in Experiment 2. (Contrast and formations in the same way as Figure 1.)

The N400 effect in native speech likely reflects increased semantic processing of blends. We predicted no N400 effect to blends in native speech because we assumed that the recognisable fragments from words/phrases in blends would be simultaneously activated, and the associated word form information would also be activated. However, our results suggest that listeners process native speech using a strong bottom-up strategy that always checks incoming word forms and actively sifts out unfitting candidates. Hence, blends still engendered an increased retrieval of lexico-semantic information in native speech.

The absence of an N400 effect to slips of the tongue in non-native speech reinforces the account suggested above based on evidence from the grammatical error condition that listeners suppress or ignore the bottom-up word form information delivered by non-native speakers. In addition, different from the classic semantic violations in Romero-Rivas et al. (2015) that were salient anomalies in their phonological forms, slips of the tongue highly resembled the intended words and consisted of fragments that might have made sense in that context. It is also possible that listeners may have suppressed or ignored these non-salient anomalies in word forms in non-native speech, as long as they couldn't directly hinder the sentence interpretation.

Interestingly, a P600 effect was evoked by slips of the tongue in both native as well as in non-native speech, whereas in native speech only grammatical errors elicited a P600 effect. These results indicate that listeners reduce their efforts in integrating incoming speech only when the speech errors encountered had been expected, for example, grammatical agreement errors that are stereotypically associated with non-native speakers. In contrast, slips of the tongue or semantic blends, in particular, are much less associated with any particular speaker identity and, thus, elicited similar P600 effects in native and non-native speech.

In sentence correctness judgements, there was no difference in the accuracy between L1 and L2 speech with blends, whereas participants performed better in detecting grammatical errors in native as compared to non-native speech. Listeners' competence of judging the correctness of L2 speech seems to be correlated with the presence and size of a P600 effect. It appears that listeners not only avoided repairing the grammatical errors in non-native speech (no P600 effect), but they were also less able to detect the errors, even in a task that strongly demanded attention to grammaticality.

Future studies should examine whether the present results can be generalised to other categories of slips of the tongue. Depending on the locus of failure within the speech production process, there might be differences in their perception.

Task-sensitivity of P600 and N400

Regarding our question about the task-sensitivity of the P600 and N400 components, our results are compatible with the previous literature that the P600 effect is bigger in direct than indirect tasks. During sentence correctness judgements, the P600 component increased robustly in its amplitude in both error conditions relative to the probe verification task. In contrast, the N400 was relatively unaffected by the task (please note that the stimuli of slips of the tongue were more homogeneous in Experiment 2 than 1).

The results could indicate that the retrieval of lexico-semantic information in sentence interpretation (N400) is relatively task-insensitive and automatic, while the integration in utterances (P600) depends strongly on where the attention is directed to under a certain communicative situation.

Effect of experience

Interestingly, the P600 effects to both error types were affected by the short-term experience in both experiments irrespective of accent. Different from Hanulíková et al. (2012) that the P600 to native grammatical errors decreased in the second half of their experiment, the P600 effect to both error types in the current study grew in Block 2. The N400 effect to blends also showed a similar dependency on experience in Experiment 2. This experience effect may be based on the repetition of our sentences in Block 2 that possibly primed some of the sentences for both speaker conditions. Accumulating experience with erroneous sentences (grammatical errors or slips of the tongue) could also have caused a more conscious attempt at retrieval and integration. The current results did not show any influence of short-term experience with a non-native accented speech on its perception.

Conclusions

In two ERP experiments, we examined how grammatical agreement violations and slips of the tongue are perceived in continuous speech, and whether native or non-native speaker identities, based on information derived from facial appearance and accent, affect the processing of different error types. We found evidence indicating different processing strategies for native and non-native speech. For grammatical agreement violations, the P600 effect was elicited only by native speech, possibly reflecting a reinterpretation process. Listeners seemed to not integrate expected error types (grammatical errors) for non-native speech. Slips of the tongue in native speech elicited N400 and P600 effects, whereas slips of the tongue in non-native speech engendered only a P600 effect, indicating that listeners pay less attention to word forms and make less effort to retrieve lexico-semantic information in non-native speech perception. We also found that short-term experience with speech errors resulted in more salient P600 effects. In addition, together the two experiments provide further evidence about the considerable task-sensitivity of P600-like components in processing speech errors and the relative automaticity of the N400 effect.

Note

1. We also conducted Cluster-based permutation tests (CBPTs) (Maris & Oostenveld, 2007) between the erroneous condition of a given error type (either slips of the tongue or grammatical agreement violations) and the corresponding correct condition to determine the time course and spatial distribution of group-level effects. Results of the CBPTs of Experiment 1 and 2 can be found in Appendix D.

Acknowledgements

We thank Antje Lorenz, Lana Rohr, Martin Maier and Julia Baum for their help with sentence construction and Luisa Balzus for her help in recording the material. We also want to thank Guido Kiecker and Gianpiero Tari for technical support, Anna Kuhlen, Antje Lorenz and the reviewers for their valuable comments and advices. The experiments used the CAS-PEAL-R1 face database collected under the sponsor of the Chinese National Hi-Tech Program and ISVISION Tech. Co. Ltd.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the China Scholarship Council (CSC) [grant number 201606230215]. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ORCID

Jue Xu http://orcid.org/0000-0003-3933-0253
Rasha Abdel Rahman http://orcid.org/0000-0002-8438-1570
Werner Sommer http://orcid.org/0000-0001-5266-3445

References

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentais, prosody, and syllable structure. *Language Learning*, *42*(4), 529–555.

Barber, H., & Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: An ERP comparison. *Journal of Cognitive Neuroscience*, *17*(1), 137–153.

Berg, P., & Scherg, M. (1994). A multiple source approach to the correction of eye artifacts. *Electroencephalography and Clinical Neurophysiology*, *90*(3), 229–241.

Brouwer, H., Crocker, M. W., Venhuizen, N. J., & Hoeks, J. C. (2017). A neurocomputational model of the N400 and the P600 in language processing. *Cognitive Science*, *41*(S6), 1318–1352.

Chen, L., Shu, H., Liu, Y., Zhao, J., & Li, P. (2007). ERP signatures of subject–verb agreement in L2 learning. *Bilingualism: Language and Cognition*, *10*(2), 161–174.

Connolly, J. F., & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience*, *6*(3), 256–266.

Cristia, A., Seidl, A., Vaughn, C., Schmale, R., Bradlow, A., & Floccia, C. (2012). Linguistic processing of accented speech across the lifespan. *Frontiers in Psychology*, *3*, 479.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21.

Ebner, N. C., Riediger, M., & Lindenberger, U. (2010). FACES – a database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior Research Methods*, *42*(1), 351–362.

Franceschina, F. (2005). *Fossilized second language grammars: The acquisition of grammatical gender (Vol. 38)*. Amsterdam: John Benjamins Publishing.

Fridlund, A. J., & Cacioppo, J. T. (1986). Guidelines for human electromyographic research. *Psychophysiology*, *23*(5), 567–589.

Gao, W., Cao, B., Shan, S., Chen, X., Zhou, D., Zhang, X., & Zhao, D. (2008). The CAS-PEAL large-scale Chinese face database and baseline evaluations. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, *38*(1), 149–161.

Grey, S., & van Hell, J. G. (2017). Foreign-accented speaker identity affects neural correlates of language comprehension. *Journal of Neurolinguistics*, *42*, 93–108.

Gunter, T. C., & Friederici, A. D. (1999). Concerning the automaticity of syntactic processing. *Psychophysiology*, *36*(1), 126–137.

Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience*, *15*(6), 883–899.

Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech compared to reading: The P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia*, *38*(11), 1531–1549.

Hahne, A., & Friederici, A. D. (1999). Electrophysiological Evidence for Two Steps in Syntactic Analysis: Early Automatic and Late Controlled Processes. *Journal of Cognitive Neuroscience*, *11*(2), 194–205. doi:10.1162/089892999563328

Hanulíková, A., van Alphen, P. M., van Goch, M. M., & Weber, A. (2012). When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience*, *24*(4), 878–887.

Heister, J., Würzner, K.-M., Bubenzer, J., Pohl, E., Hanneforth, T., Geyken, A., & Kliegl, R. (2011). dlexDB – eine lexikalische Datenbank für die psychologische und linguistische Forschung [dlexDB – a lexical data bank for psychological and linguistic researches]. *Psychologische Rundschau*, *62*, 10–20.

Koester, D., Gunter, T. C., Wagner, S., & Friederici, A. D. (2004). Morphosyntax, prosody, and linking elements: The auditory processing of German nominal compounds. *Journal of Cognitive Neuroscience*, *16*(9), 1647–1668.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, *1146*, 23–49.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud faces database. *Cognition and Emotion*, *24*(8), 1377–1388.

Lattner, S., & Friederici, A. D. (2003). Talker's voice and gender stereotype in human auditory sentence processing–evidence from event-related brain potentials. *Neuroscience Letters*, *339*(3), 191–194.

Leuninger, H. (1996). *Danke und Tschüss fürs Mitnehmen: gesammelte Versprecher und eine kleine Theorie ihrer Korrekturen [Thanks and bye to go: collected slips of the tongue and a small theory about their corrections]*. Munich: Ammann.

Leuninger, H. (1999). *Reden ist Schweigen, Silber ist Gold: gesammelte Versprecher [Talk is silence, silver is golden: Collected slips of the tongue]*. Leipzig: Dt. Taschenbuch-Verlag.

Lindenberger, U., Ebner, N., & Riediger, M. (2005–2007). FACES collection, from Max Planck Institute for Human Development, Center for Lifespan Psychology, Berlin, Germany. Retrieved from http://faces.mpib-berlin.mpg.de/album/escidoc:57488

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190.
Martín-Loeches, M., Nigbur, R., Casado, P., Hohlfeld, A., & Sommer, W. (2006). Semantics prevalence over syntax during sentence processing: A brain potential study of noun-adjective agreement in Spanish. *Brain Research*, *1093*(1), 178–189.
Meringer, R., & Mayer, K. (1895). *Versprechen und Verlesen: eine Psychologisch-linguistische Studie [slips of the tongue and Misreading: A psychological linguistic study]*. Stuttgart: Göschensche Verlagsbuchhandlung.
Molinaro, N., Barber, H. A., & Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex*, *47*(8), 908–930.
Molinaro, N., Vespignani, F., & Job, R. (2008). A deeper reanalysis of a superficial feature: An ERP study on agreement violations. *Brain Research*, *1228*, 161–176.
Mueller, J. L., Oberecker, R., & Friederici, A. D. (2009). Syntactic learning by mere exposure-an ERP study in adult learners. *BMC Neuroscience*, *10*(1), 89.
Munro, M. J. (2003). A primer on accent discrimination in the Canadian context. *TESL Canada Journal*, *20*(2), 38–51.
Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113.
Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). Fieldtrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 1–9.
Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329–347.
Rabovsky, M., Hansen, S. S., & McClelland, J. L. (2018). Modelling the N400 brain potential as change in a probabilistic representation of meaning. *Nature Human Behaviour*, *2*(9), 693–705. doi:10.1038/s41562-018-0406-4
Roehm, D., Bornkessel, I., Haider, H., & Schlesewsky, M. (2005). When case meets agreement: Event-related potential effects for morphology-based conflict resolution in human language comprehension. *NeuroReport*, *16*(8), 875–878.
Romero-Rivas, C., Martin, C. D., & Costa, A. (2015). Processing changes when listening to foreign-accented speech. *Frontiers in Human Neuroscience*, *9*, 167.
Romero-Rivas, C., Martin, C. D., & Costa, A. (2016). Foreign-accented speech modulates linguistic anticipatory processes. *Neuropsychologia*, *85*, 245–255.
Rossi, S., Gugler, M. F., Hahne, A., & Friederici, A. D. (2005). When word category information encounters morphosyntax: An ERP study. *Neuroscience Letters*, *384*(3), 228–233.
Sabourin, L., Stowe, L. A., & De Haan, G. J. (2006). Transfer effects in learning a second language grammatical gender system. *Second Language Research*, *22*(1), 1–29.
Schacht, A., Sommer, W., Shmuilovich, O., Martinez, P. C., & Martin-Loeches, M. (2014). Differential task effects on N400 and P600 elicited by semantic and syntactic violations. *PLoS One*, *9*(3), e91226.
van Berkum, J. J. (2004). Sentence comprehension in a wider discourse: Can we use ERPs to keep track of things. In M. Carreiras & C. Clifton, Jr. (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERPs and beyond* (pp. 229–270). New York: Psychology Press.
van Berkum, J. J., van den Brink, D., Tesink, C. M., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, *20*(4), 580–591.
van den Brink, D., Brown, C. M., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience*, *13*(7), 967–985.
van Herten, M., Kolk, H. H., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*, *22*(2), 241–255. doi:10.1016/j.cogbrainres.2004.09.002
van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(2), 394–417.
Wietzel-Winkler, S. (2017, May). *Unsere Versprecher & Verhaspler [Our slips of the tongue]*. Retrieved from http://www.zur-sprache-kommen.de/?Unsere_Versprecher

Original Research Article II

II. Sequential Adaptation Effects Reveal Proactive Control in Processing Spoken Sentences: Evidence from Event-related Potentials

Jue Xu, Rasha Abdel Rahman, & Werner Sommer

Reference:

Xu, J., Abdel Rahman, R., & Sommer, W. (2021). Sequential Adaptation Effects Reveal Proactive Control in Processing Spoken Sentences: Evidence from Event-related Potentials. *Brain and Language*, 214, 104904. doi: 10.1016/j.bandl.2020.104904

Brain & Language 214 (2021) 104904

Contents lists available at ScienceDirect

Brain and Language

journal homepage: www.elsevier.com/locate/b&l

Sequential adaptation effects reveal proactive control in processing spoken sentences: Evidence from event-related potentials

Jue Xu *, Rasha Abdel Rahman, Werner Sommer

Institut für Psychologie, Humboldt-Universität zu Berlin, Rudower Chaussee 18, 12489 Berlin, Germany

ARTICLE INFO

Keywords:
Cognitive control
P600
Proactive control
Sequential effects
Speech errors

ABSTRACT

How domain-general cognitive control is engaged in language processing remains debated. We address how linguistic processes are monitored and regulated by analyzing the effects of previous-trial sentence correctness on the P600 component of the event-related potential (ERP) in the current-trial. In data from a previous experiment about processing spoken sentences, P600 amplitudes to both correct and incorrect words in current sentences were smaller after incorrect as compared to correct previous sentences. Therefore, the detection of speech errors may initiate sustained proactive control over the monitoring demands for upcoming sentences. No sequential adaptation was found in the difference between P600 amplitudes to incorrect and correct current conditions. We propose that the P600 reflects the reactive reanalysis of speech processing and/or the resolution of linguistic conflicts, but is also sensitive to proactive speech monitoring, an important aspect of cognitive control.

1. Introduction

Cognitive or executive control must be exerted in many situations and tasks (Diamond, 2013) and may also be necessary for goal- and context-appropriate language processing and comprehension (see Key-DeLyria & Altmann, 2016, for a review). In typical non-linguistic cognitive control paradigms like Stroop, Flanker, and Simon tasks, conflicts arise between task-relevant and task-irrelevant stimulus properties and/or stimulus–response associations (Gratton, Cooper, Fabiani, Carter, & Karayanidis, 2018). The detection of conflict initiates sustained cognitive control mechanisms, which may enhance the resolution of subsequent conflict(s), yielding "conflict adaptation" effects (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Gratton, Coles, & Donchin, 1992; Gratton et al., 2018). Participants tend to respond more quickly and more accurately to incongruent trials after incongruent rather than congruent trials (Hommel, Proctor, & Vu, 2004; Kerns et al., 2004; Stürmer, Leuthold, Soetens, Schröter, & Sommer, 2002). To explain conflict adaptation effects, Braver, Gray, and Burgess (2007) proposed a dual mechanism model of proactive and reactive control (DMC): If the processing of a stimulus produces a conflict, reactive control is triggered in order to resolve the conflict, but it may also trigger proactive control, maintaining task- or context-relevant information in memory, anticipating and preventing upcoming conflicts (Braver et al., 2007).

Many studies indicate the involvement of executive control in language processing. For example, January, Trueswell, and Thompson-Schill (2009) found overlapping BOLD activation to syntactic ambiguity and Stroop-like incongruency. Conflict adaptation effects may transfer across tasks and domains: sustained cognitive control initiated by previous conflicts (i.e., incongruent Stroop trials) may facilitate resolving ambiguities in subsequent sentences (Novick, Hussey, Teubner-Rhodes, Harbison, & Bunting, 2014; Hsu & Novick, 2016), demonstrating that resolving sentence ambiguity may be subject to proactive cognitive control.

Conversely, conflict detection in the syntactic domain seems to facilitate conflict resolution in the Stroop task (Kan et al., 2013; Vuong & Martin, 2014). In sentence processing, syntactic ambiguities or anomalies may induce conflicts (i.e., mismatches) either between the expected and the actually encountered linguistic representations or between different response alternatives (van de Meerendonk, Kolk, Chwilla, & Vissers, 2009). In a word-by-word reading paradigm alternating with Stroop trials, Kan et al. (2013) found better performance in incongruent Stroop trials after incongruent (garden-path) sentences rather than congruent (unambiguous) sentences. These results indicate that encountering conflicts in sentences may initiate sustained proactive control.

The findings above reveal cross-domain conflict adaptation effects between syntactic ambiguity and cognitive conflicts. The DMC has been proposed to also explain cognitive control in language processes (Berry,

* Corresponding author.
E-mail addresses: xujue@hu-berlin.de (J. Xu), rasha.abdel.rahman@hu-berlin.de (R. Abdel Rahman), werner.sommer@cms.hu-berlin.de (W. Sommer).

https://doi.org/10.1016/j.bandl.2020.104904
Received 7 July 2020; Received in revised form 15 November 2020; Accepted 22 December 2020
Available online 14 January 2021

2016; Columbus et al., 2015). However, the domain-generality of the proposed cognitive control in language processing remains to be determined (see Fedorenko, 2014, for a review). Previous findings suggest either a domain-general cognitive control system that is shared across syntactic and non-syntactic domains or a domain-specific cognitive control system for syntactic and verbal conflicts (Kan et al., 2013; Novick et al., 2014; Hsu & Novick, 2016; Vuong & Martin, 2014). In any case, one should expect conflict adaptation within the syntactic domain; more specifically, the detection of syntactic errors or anomalies might initiate control processes that proactively influence the processing of the following sentence(s). To the best of our knowledge, this hypothesis has yet to be tested. Here we explored this idea by investigating conflict adaptation or sequential effects in electrophysiological indices of language processing in correct and incorrect sentences.

EEG studies reported that deviations from linguistic well-formedness, such as garden-path sentences (Kaan & Swaab, 2003; Osterhout & Holcomb, 1992), syntactic violations (Friederici, Pfeifer, & Hahne, 1993; Hagoort, Brown, & Groothusen, 1993), semantic and orthographic anomalies (van Herten, Kolk, & Chwilla, 2005; Vissers, Chwilla, & Kolk, 2006), typically elicit a P600 component in event-related potentials (ERPs). P600 effects are maximal at centro-parietal scalp sites, typically start around 500 ms after word onset and may extend to one second or more. The P600 effect is small or even absent when the proportion of linguistic errors within a block or experiment is high (e.g., Coulson, King, & Kutas, 1998; Hahne & Friederici, 1999; Xu, Abdel Rahman, & Sommer, 2019). The P600 effect is also larger for direct tasks, like correctness judgments, than for indirect tasks, like probe verifications (Gunter & Friederici, 1999; Schacht, Sommer, Shmuilovich, Martienz, & Martin-Loeches, 2014; Xu et al., 2019).

The consistency of the P600 effect across various kinds of speech errors and ambiguities and its susceptibility to task relevance and probability suggest a close interaction between the executive system and the language system (e.g., Coulson et al., 1998; Kolk & Chwilla, 2007). Cognitive control is theoretically linked to the P600 by the monitoring account of language perception (van de Meerendonk et al., 2009). Monitoring is an evaluative component in cognitive control that keeps contextual information activated and entails the detection of conflicts and the triggering of compensatory adjustments in control (Botvinick et al., 2001). The monitoring account of language perception proposed that a mismatch between competing linguistic representations or uncertainty in how to respond may be processed like a conflict. Detecting such conflicts may trigger a reanalysis of the sentence for processing errors, manifested in the P600 effect (van de Meerendonk et al., 2009). Considering the reports of cross-domain conflict adaptation effects reviewed above, the monitoring account would predict that conflicts in the syntactic domain in a given sentence should modulate proactive control over the processing of the following sentence, causing sequential adaptation effects reflected in P600 modulations.

According to the DMC, classic P600 effects – i.e., larger P600 amplitudes to linguistic errors or ambiguities relative to well-formedness – can be viewed as resolution mechanisms triggered reactively by detecting a conflict. Conflict adaptation in non-linguistic paradigms typically manifests as interaction between previous and current resolution costs in performance (Botvinick et al., 2001). If this holds true also for linguistic contexts, experiment-wide diminished P600 effects to speech errors under conditions of high syntactic error proportion and in indirect tasks (e.g., Xu et al., 2019) may reflect experiment-wide proactive control; context information such as error proportion and task demands is updated and maintained in working memory to facilitate the resolution of upcoming conflicts. The experiment-wide diminished P600 effects might be the consequence of accumulated sequential adaptation effects within the experiment. If so, one might expect smaller P600 effects after speech errors in sentences following ill-formed relative to well-formed sentences – similar to sequential effects observed in cognitive conflict tasks (e.g., Stürmer et al., 2002).

Although the reasoning above might suggest analogous sequential adaptation effects for linguistic and non-linguistic tasks, it is prudent to consider an alternative prediction. After all, the nature of conflicts in cognitive paradigms, such as Stroop, or Simon tasks, and in sentence processing is very different. In the former tasks conflicts may arise between different stimulus properties and/or stimulus–response associations, whereas conflicts during sentence processing may occur between perceivers' subjectively expected linguistic input and what is actually encountered (van de Meerendonk et al., 2009). Thus, anticipation and conflict resolution in Stroop tasks are based on biasing attention towards or away from stimulus properties and/or stimulus–response associations (Gratton et al., 2018). In contrast, expectations in speech perception can be formed on the basis of established general rules and different kinds of linguistic and social contexts (e.g., Hanulíková, van Alphen, van Goch, & Weber, 2012; van Berkum, van den Brink, Tesink, Kos, & Hagoort, 2008; Xu et al., 2019). Conflict resolution or general reanalysis of speech processing may engage the integration of newly encountered linguistic inputs into existing mental representations and/or reinterpretation of sentences through syntactic analysis. From a linguistic perspective, the semantic integration and syntactic processing in the current sentence should bear little or no influence from detecting speech errors in the preceding one, i.e., from instantaneously increased expectation for errors. Rather, different processes in cognitive control (e.g., attention, working memory, inhibition, etc.) (Diamond, 2013; Miyake et al. 2000) may affect how speech processing is monitored in general. In other words, proactive control mechanisms might be independent of and additive to the resolution processes for conflicts encountered in speech processing. The detection of linguistic conflicts may increase error expectation for the following sentence and facilitate the monitoring of speech processing through, for example, proactive maintenance of P600-related resources – meaning that fewer resources would have to be additionally recruited for the next sentence, resulting in smaller P600 amplitudes in the upcoming sentence. Additionally, as revealed by its sensitivity to linguistic error probability, P600 amplitudes generally decrease under higher expectation for errors in both correct and incorrect linguistic inputs (e.g., Coulson et al., 1998). Therefore, detecting a speech error in a previous sentence may diminish P600 amplitudes to critical words in a current sentence, independent of its linguistic correctness, at variance with the sequential effects observed in cognitive conflict tasks (Stürmer et al., 2002).

Improvements in accuracy after preceding incongruent as compared to congruent trials can be found in cognitive conflict paradigms (e.g., the Flanker task: Gratton et al., 1992; the Simon task: Stürmer et al., 2002) and in cross-domain conflict adaptation paradigms using garden-path sentences (e.g., Hsu & Novick, 2016; Kan et al., 2013). Similarly, responses in judging the current-sentence correctness may be more accurate after having detected an error in the previous sentence, as compared to having detected no errors. Thus we also investigated whether error detection on Sentence N-1 had affected judgment accuracies on Sentence N, which might provide additional behavioral support for sequential modulations of speech processing strategies.

To conclude, the assumptions that drive this article are first, that a variety of cognitive functions are used during sentence processing and comprehension. Second, processing syntactic ambiguity may initiate and be influenced by sustained cognitive control during sentence processing and comprehension. Finally, P600, reflecting monitoring and conflict resolution of speech processing, is governed by reactive and proactive executive control in a task- and context-relevant manner. In light of these assumptions, we examined whether the detection of syntactic errors in preceding sentences affects current-trial P600 in correct and incorrect sentences. We expected that after detecting an error in a preceding sentence, P600 amplitudes in the current sentence would decrease for both correct and incorrect words. To investigate this prediction, we reanalysed an existing EEG data set (Xu et al., 2019, Experiment 2).

2. Methods

2.1. Previous study

In Xu et al. (2019, Exp. 2), a total of 24 participants was tested. German sentences that were either well-formed (correct) or contained a speech error (grammatical agreement violation or semantic blend), spoken in either native- or foreign-accented voices, were presented auditorily, randomly interspersed, preceded (and accompanied) by portraits of European or Asian faces, respectively. Each test session consisted of 360 sentences. The proportion of incorrect sentences within and across the native and non-native speaker groups was 50%. With equal probabilities, correct and incorrect sentences could be preceded by either a correct or incorrect sentence. Examples for each type of sentence are provided in Table 1.

Xu et al. (2019) conducted list-wide analyses on each error type and speaker type and revealed that P600 effects were only evoked by grammatical errors in native but not in non-native speech, and that semantic blends in both native and non-native speech elicited P600 effects. Sequential effects had not been investigated.

2.2. Data analysis

All data from the original study were included and pre-processed in the same way in the present reanalysis. The reader interested in details is kindly referred to this publication (Xu et al., 2019). There were 45 well-formed filler sentences in each session, which were correct counterparts of phonological blends from Experiment 1 of Xu et al. (2019). Therefore, ERPs for the fillers in the present analysis were time-locked to the phonological divergence points between the two conditions. All trials were included in the present analysis in terms of their effects on subsequent critical sentences and in terms as sentences following correct or incorrect predecessors.

The experimental task in Experiment 2 (Xu et al., 2019) was to judge the correctness of each sentence immediately after its presentation. No feedback on the judgment was given during the test. These correctness judgments were used as subjective measures of whether an error was detected or not on Trial N-1, because studies of cross-domain sequential effects (e.g., Kan et al., 2013) indicate that only readers' actual detection of conflicts rather than objective correctness will trigger proactive control over upcoming conflict(s). If the previous sentence was judged as correct, it was taken to reflect that no error had been detected; if the previous sentence was judged as incorrect, it was taken to reflect that the participant had detected an "error" in the previous sentence. For current sentences, conforming to common practice in assessing P600 effects to linguistic inputs and same as in the previous study (Xu et al., 2019), current-trial sentence correctness was determined by the objective correctness, i.e., actual correctness of the sentence. Trials with inaccurate responses were also included in the analysis.

In the analysis, we collapsed the current and previous sentence correctness conditions across native and non-native speaker types and across the two error types. Proactive mechanisms were assumed to be triggered by error detection and to affect general error expectation, regardless of the type of error or speaker. Thus the different conditions were collapsed for previous sentence condition. As proposed above, P600 effects to speech errors are taken to reflect reactive mechanisms for conflict resolution. In current sentences, influences from current speaker or error type should co-exist with any proactive effects triggered already in previous sentences. Moreover, proactive mechanisms were assumed to affect P600 for both correct and incorrect linguistic inputs. As revealed by the original study, only non-native grammatical errors (less than 12.8% of all sentences) had no P600 effects compared to their correct counterparts. Therefore the current speaker or error type should not need to be considered in studying proactive mechanisms in the P600 in the present analysis.

We distinguished four correctness sequences of two consecutive sentences: correct-incorrect (judged-as-correct previous sentence - incorrect current sentence), correct-correct, incorrect-incorrect, and incorrect-correct. After considering previous sentence correctness based on subjective judgments and after automatic artefact rejection in EEG processing, the average trial numbers for the sequences correct-incorrect, correct-correct, incorrect-incorrect, and incorrect-correct were 103 (SD = 16), 101 (SD = 17), 76 (SD = 12), and 78 (SD = 13), respectively. Accuracies of correctness judgments were calculated separately for sentences with judged-as-correct and judged-as-incorrect preceding trials and entered into a one-way ANOVA with factor previous correctness (judged as correct vs. incorrect).

As in Xu et al. (2019), for the P600 analyses, mean amplitudes across the 25 electrodes in the ROI (C1/2, C3/4, CP1/2, CP3/4, CP5/6, P3/P4, P5/6, PO3/4, PO7/8, O1/2, Cz, CPz, Pz, POz, Oz) were calculated between 600 and 1200 ms. To assess sequential effects of sentence correctness on the P600, amplitudes were entered into a repeated measures ANOVA with factors current correctness (correct vs. incorrect) and previous correctness (judged as correct vs. incorrect).

3. Results

Fig. 1 shows the P600 amplitudes to current trials in the four correctness sequences with regard to previous trials. ANOVA on the P600 revealed a significant main effect of previous correctness ($F(1, 23) = 10.72$, $p = .003$, $\eta_p^2 = 0.318$): The P600 amplitude was smaller after previous sentences where errors had been detected as compared to subjectively correct sentences. The ANOVA also revealed a significant main effect of current correctness ($F(1, 23) = 8.91$, $p = .007$, $\eta_p^2 = 0.279$): The P600 amplitude of words in the current sentences was larger for incorrect compared to correct current sentences, reflecting classic P600 effects. Current and previous correctness did not interact ($F < 1$).

Participants showed very similar accuracies of sentence correctness judgments on current trials after previous trials where speech errors had been detected (M = 82.90%, SD = 4.99%) as compared to subjectively correct sentences (M = 82.74%, SD = 4.03%). A one-way ANOVA of previous correctness failed significance by a wide margin ($F < 1$).

4. Discussion

This reanalysis of an existing data set found sequential adaptation effects in processing spoken sentences. The P600 amplitude to critical words in current sentences was smaller when the immediately preceding sentence had been judged as incorrect rather than correct. This novel result was independent of and additive to the well-known immediate effect of syntactic sentence correctness on the P600.

Based on the DMC model of proactive and reactive cognitive control (Braver et al. 2007), developed for conflict tasks, and on the monitoring theory of P600 (van de Meerendonk et al., 2009), we proposed that the P600 component elicited in linguistic contexts is susceptible to influences from both pro- and reactive control. More specifically, we

Table 1
Sentence Examples.

a. well-formed	Ich hätte gerne einen$_{[m]}$ Mocca$_{[m]}$ mit Sahne.
(English translation:	I would like a$_{[m]}$ mocha$_{[m]}$ with cream.)
b. with a grammatical agreement violation	Ich hätte gerne eine$_{[f]}$ Mocca$_{[m]}$ mit Sahne.
c. with a semantic blend	Ich hätte gerne einen Moccolade* mit Sahne. * Moccolade = Mocca (mocha) + Schokolade (chocolate)

Notes: Sentence examples are taken from Appendix A of Xu et al. (2019). Single-underlined word is trigger for the condition grammatical agreement violations. Wavy-underlined word is trigger for the condition semantic blends. Grammatical gender (m = masculine, f = feminine) refers to the gender of this noun if subscripted under a noun; otherwise, it refers to the correct gender that the determiner should lead.

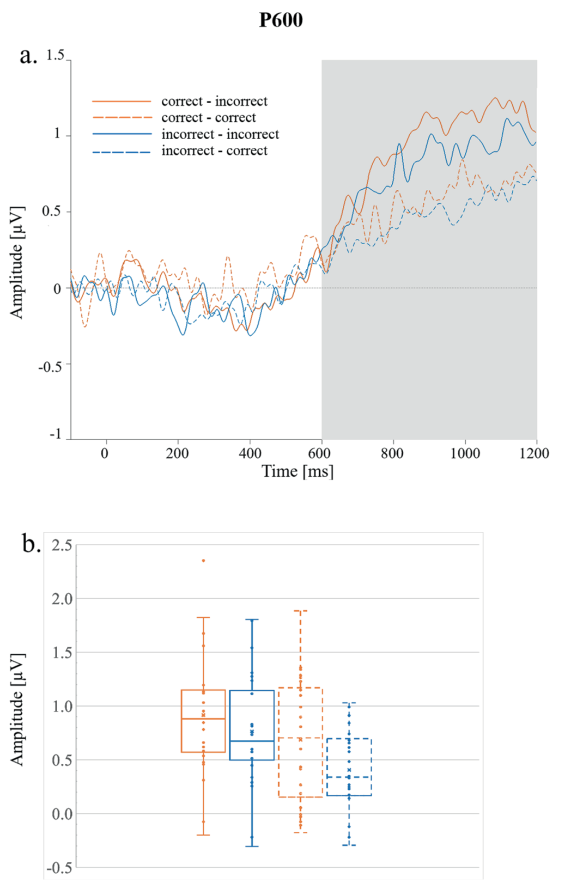


Fig. 1. P600 Amplitudes in the four Correctness Sequence Conditions. ERPs represent grand means (N = 24) in P600-ROI averaged separately for four correctness sequence conditions (previous sentence – current sentence) in Panel a. Positive is plotted upward. Time window for the P600 effect is shaded. Averaged P600-ROI amplitudes of all 24 participants for four correctness sequence conditions are displayed in the boxplot in Panel b.

suggest that the P600 effect after detecting speech errors or complexities as compared to well-formed sentences reflects reactive conflict resolution. In addition, the detection of speech errors should initiate sustained proactive control over speech monitoring for the following sentence, reflected by decreased P600 amplitudes when encountering subjectively incorrect as compared to correct sentences.

The main function of proactive control in DMC is anticipating conflicts and maintaining critical information in memory (Braver et al., 2007). The sequential adaptation in P600 appears to be the consequence of detecting a linguistic error. Detecting a linguistic error possibly leads to higher expectation for linguistic conflicts in the upcoming sentence, diminishing P600 amplitudes to both correct and incorrect words in current sentences. Given that there is little or no syntactic conflict in correct current sentences, the decreased P600 amplitudes in these sentences are most naturally attributed to a shared general mechanism, for example monitoring, rather than specific error-related mechanisms. Detecting an error and increased error expectation can be considered as a kind of information. Proactive control updates and maintains this information in working memory, and neural resources for monitoring speech processing are proactively maintained in preparation of handling the upcoming speech input. Therefore, we propose that proactive control may regulate the deployment of monitoring speech processing through adjusting the expectation for upcoming sentence input.

This decrease in neural responses across consecutive sentences might be similar or related to the repetition suppression phenomenon, which refers to a reduction in neural activity induced by stimulus repetition, one of the most robust experience-related cortical dynamics (see Grill-Spector, Henson, & Martin, 2006 for a review). Future research might investigate parallels between the sequential effects in P600 observed here and other ERP components, for example, the P300 regarding the P600-as-P3 account (Coulson et al., 1998; Sassenhagen, Schlesewsky, & Bornkessel-Schlesewsky, 2014).

The current reanalysis did not reveal an influence of previous-trial correctness on current-trial P600 effects or on the accuracy of current-trial judgments, i.e., previous-sentence errors did not modulate the neural manifestations of resolving speech error(s), as might have been expected from findings in cognitive conflict tasks. However, the additivity of preceding and current correctness does not conclusively prove the independence of pro- and reactive control in speech processing. Fig. 1 shows a numerically larger difference of P600 amplitudes after incorrect preceding sentences, indicating that the presence of an interaction might depend on testing power and more homogeneous materials (Xu et al., 2019; see also the additional analyses in the supplement). Nevertheless, the direction of the trend in Fig. 1 is opposite to the experiment-wide reduction of the P600 effect under high error proportions, at variance with the assumption about the experiment-wide reduced P600 effect being the consequence of cumulated sequential adaptation effects in the experiment. However, we believe that proactive and reactive control mechanisms in language processing should at least be interactive in the way that reactive control sends critical information about detecting linguistic conflicts to the monitoring device, enabling the sequential proactive adaptation observed here.

As explained in the introduction, proactive control involved here must at least apply beyond the syntactic domain (e.g., Kan et al., 2013; Novick et al., 2014; Hsu & Novick, 2016). Since cognitive control consists in many different processes (e.g., attention, working memory, inhibition, etc.) (Diamond, 2013; Miyake et al. 2000), it is conceivable that proactive control consists in one process (e.g., monitoring), whereas reactive control is either a different general process (e.g., working memory) or more language-specific, involving the explanations suggested for the P600. In any case, the current findings do not reconcile with a single process deployed for both pro- and reactive control in processing spoken sentences. It is conceivable that proactive control might also affect processing strategies at other levels such as the amount of resources invested into reactive control or conflict resolution based on accumulated reactive feedbacks about high ratios of errors or conflicts. Similar dual processing mechanisms as outlined in the DMC model (Braver et al., 2007) seem to hold true also for processing spoken sentences.

Since P600 is sensitive to task demands (e.g., Schacht et al., 2014; Xu et al., 2019), there have also been concerns that conclusions based on findings about P600 under an artificial experimental task (i.e., acceptability judgments) do not apply to language processing under presumably more natural conditions (e.g., passive reading/listening): When the task is to find speech errors, the context of task places a variety of cognitive demands on participants apart from language comprehension (e.g., Campbell & Tyler, 2018). However, cognitive functions such as attention and memory that are demanded by the judgment task also contribute to linguistic processes – perhaps not in pure syntactic processing, but possibly for processing semantics and pragmatics in the wider language system. Besides, previous cross-domain conflict adaptation effects were found without requesting acceptability judgments of the garden-path sentences but rather comprehending their meanings (e.g., Novick et al., 2014; Hsu & Novick, 2016). It is thus implausible that findings under acceptability judgment tasks are merely artefacts of the task.

5. Conclusions

In the present study we found decreased P600 amplitudes in current sentences after detecting an error in previous sentences, indicating a previously unreported sequential adaptation effect across sentences, possibly reflecting adaptation in the monitoring demands for future events. This novel finding has important implications for understanding when and how cognitive control is engaged during language comprehension. Similar dual processing mechanisms as outlined in the DMC model (Braver et al., 2007) seem to hold true also for speech processing. However, the nature of proactive control processes in linguistic contexts seems to be different and more general than in cognitive conflict tasks.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank the reviewers for their valuable comments.

Funding

This work was supported by the China Scholarship Council (CSC) under Grant [201606230215] to J.X. The funder had no role in data analysis, decision to publish, or preparation of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bandl.2020.104904.

References

Berry, G. M. (2016). Processing linguistic variation through dual mechanisms of cognitive control. *Linguistics Vanguard, 2*(s1). https://doi.org/10.1515/lingvan-2016-0033.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*(3), 624. https://doi.org/10.1030//0033-295X.108.3.624.

Braver, T. S., Gray, J. R., & Burgess, G. C. (2007). Explaining the many varieties of working memory variation: Dual mechanisms of cognitive control. In A. R. A. Conway, C. Jarrold, M. J. Kane (Eds.) & A. Miyake & J. N. Towse (Ed.), Variation in working memory (pp. 76–106). New York, US: Oxford University Press.

Campbell, K. L., & Tyler, L. K. (2018). Language-related domain-specific and domain-general systems in the human brain. *Current Opinion in Behavioral Sciences, 21*, 132–137. https://doi.org/10.1016/j.cobeha.2018.04.008.

Columbus, G., Sheikh, N. A., Côté-Lecaldare, M., Häuser, K., Baum, S. R., & Titone, D. (2015). Individual differences in executive control relate to metaphor processing: An eye movement study of sentence reading. *Frontiers in Human Neuroscience, 8*, 1057. https://doi.org/10.3389/fnhum.2014.01057.

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes, 13*(1), 21–58. https://doi.org/10.1080/016909698386582.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*, 135–168. https://doi.org/10.1146/annurev-psych-113011-143750.

Fedorenko, E. (2014). The role of domain-general cognitive control in language comprehension. *Frontiers in Psychology, 5*, 335. https://doi.org/10.3389/fpsyg.2014.00335.

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive Brain Research, 1*(3), 183–192. https://doi.org/10.1016/0926-6410(93)90026-2.

Gratton, G., Coles, M. G., & Donchin, E. (1992). Optimizing the use of information: Strategic control of activation of responses. *Journal of Experimental Psychology: General, 121*(4), 480. https://doi.org/10.1037/0096-3445.121.4.480.

Gratton, G., Cooper, P., Fabiani, M., Carter, C. S., & Karayanidis, F. (2018). Dynamics of cognitive control: Theoretical bases, paradigms, and a view for the future. *Psychophysiology, 55*(3), e13016. https://doi.org/10.1111/psyp.13016.

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: Neural models of stimulus-specific effects. *Trends in Cognitive Sciences, 10*(1), 14–23. https://doi.org/10.1016/j.tics.2005.11.006.

Gunter, T. C., & Friederici, A. D. (1999). Concerning the automaticity of syntactic processing. *Psychophysiology, 36*(1), 126–137. https://doi.org/10.1017/S004857729997155X.

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes, 8*(4), 439–483. https://doi.org/10.1080/01690969308407585.

Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. *Journal of Cognitive Neuroscience, 11*(2), 194–205. https://doi.org/10.1162/089892999563328.

Hanulíková, A., van Alphen, P. M., van Goch, M. M., & Weber, A. (2012). When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience, 24*(4), 878–887. https://doi.org/10.1162/jocn_a_00103.

Hommel, B., Proctor, R. W., & Vu, K.-P.-L. (2004). A feature-integration account of sequential effects in the Simon task. *Psychological Research Psychologische Forschung, 68*(1), 1–17. https://doi.org/10.1007/s00426-003-0132-y.

Hsu, N. S., & Novick, J. M. (2016). Dynamic engagement of cognitive control modulates recovery from misinterpretation during real-time language processing. *Psychological Science, 27*(4), 572–582. https://doi.org/10.1177/0956797615625223.

January, D., Trueswell, J. C., & Thompson-Schill, S. L. (2009). Co-localization of Stroop and syntactic ambiguity resolution in Broca's area: Implications for the neural basis of sentence processing. *Journal of Cognitive Neuroscience, 21*(12), 2434–2444. https://doi.org/10.1162/jocn.2008.21179.

Kan, I. P., Teubner-Rhodes, S., Drummey, A. B., Nutile, L., Krupa, L., & Novick, J. M. (2013). To adapt or not to adapt: The question of domain-general cognitive control. *Cognition, 129*(3), 637–651. https://doi.org/10.1016/j.cognition.2013.09.001.

Kaan, E., & Swaab, T. Y. (2003). Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of Cognitive Neuroscience, 15*(1), 98–110. https://doi.org/10.1162/089892903321107855.

Kerns, J. G., Cohen, J. D., MacDonald, A. W., Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science, 303*(5660), 1023–1026. https://doi.org/10.1126/science.1089910.

Key-DeLyria, S. E., & Altmann, L. J. (2016). Executive function and ambiguous sentence comprehension. *American Journal of Speech-Language Pathology, 25*(2), 252–267. https://doi.org/10.1044/2015_AJSLP-14-0111.

Kolk, H., & Chwilla, D. (2007). Late positivities in unusual situations. *Brain and Language, 100*(3), 257–261. https://doi.org/10.1016/j.bandl.2006.07.006.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734.

Novick, J. M., Hussey, E., Teubner-Rhodes, S., Harbison, J. I., & Bunting, M. F. (2014). Clearing the garden-path: Improving sentence processing through cognitive control training. *Language, Cognition and Neuroscience, 29*(2), 186–217. https://doi.org/10.1080/01690965.2012.758297.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31*(6), 785–806. https://doi.org/10.1016/0749-596X(92)90039-Z.

Sassenhagen, J., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2014). The P600-as-P3 hypothesis revisited: Single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned. *Brain and Language, 137*, 29–39. https://doi.org/10.1016/j.bandl.2014.07.010.

Schacht, A., Sommer, W., Shmuilovich, O., Martienz, P. C., & Martin-Loeches, M. (2014). Differential task effects on N400 and P600 elicited by semantic and syntactic violations. *PLoS ONE, 9*(3), e91226. https://doi.org/10.1371/journal.pone.0091226.

Stürmer, B., Leuthold, H., Soetens, E., Schröter, H., & Sommer, W. (2002). Control over location-based response activation in the Simon task: Behavioral and electrophysiological evidence. *Journal of Experimental Psychology: Human Perception and Performance, 28*(6), 1345. https://doi.org/10.1037/0096-1523.28.6.1345.

van Berkum, J. J., van den Brink, D., Tesink, C. M., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience, 20*(4), 580–591. https://doi.org/10.1162/jocn.2008.20054.

van de Meerendonk, N., Kolk, H. H., Chwilla, D. J., & Vissers, C. T. W. (2009). Monitoring in language perception. *Language and Linguistics Compass, 3*(5), 1211–1224. https://doi.org/10.1111/j.1749-818X.2009.00163.x.

van Herten, M., Kolk, H. H., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Brain Research. Cognitive Brain Research, 22*(2), 241–255. https://doi.org/10.1016/j.cogbrainres.2004.09.002.

Vissers, C. T., Chwilla, D. J., & Kolk, H. H. (2006). Monitoring in language perception: The effect of misspellings of words in highly constrained sentences. *Brain Research, 1106*(1), 150–163. https://doi.org/10.1016/j.brainres.2006.05.012.

Vuong, L. C., & Martin, R. C. (2014). Domain-specific executive control and the revision of misinterpretations in sentence comprehension. *Language, Cognition and Neuroscience, 29*(3), 312–325. https://doi.org/10.1080/01690965.2013.836231.

Xu, J., Abdel Rahman, R., & Sommer, W. (2019). Perceived language competence modulates criteria for speech error processing: Evidence from event-related potentials. *Language, Cognition and Neuroscience, 1–14*. https://doi.org/10.1080/23273798.2018.1562558.

Original Research Articles

III. Who Speaks Next? Adaptations to Speaker Identity in Processing Spoken Sentences

Jue Xu, Rasha Abdel Rahman, & Werner Sommer

Manuscript under review for *Psychophysiology.*

Who Speaks Next? Adaptations to Speaker Identity in Processing Spoken Sentences

Jue Xu[a], Rasha Abdel Rahman[b], Werner Sommer[c]

[a, b, c] *Institut für Psychologie, Humboldt-Universität zu Berlin,*

Correspondence Address: Rudower Chaussee 18, 12489 Berlin

[a] *ORCID ID:0000-0003-3933-0253*

Telephone Number: +49-(0)30-20934904

Email Address: xujue@hu-berlin.de

[b] *ORCID ID: 0000-0002-8438-1570*

Telephone Number: +49-(0)30-20939413

Email Address: rasha.abdel.rahman@hu-berlin.de

[c] *ORCID ID: 0000-0001-5266-3445*

Telephone Number: +49-(0)30-20934886

Email Address: werner.sommer@cms.hu-berlin.de

Who Speaks Next? Adaptations to Speaker Identity in Processing Spoken Sentences

Abstract

When listening to a speaker, we need to adapt to her individual speaking characteristics, such as error proneness, accent, etc. The present study investigated two aspects of adaptation to speaker identity during processing spoken sentences in multi-speaker situations: the effect of speaker sequence across sentences and the effect of learning speaker-specific error probability. Spoken sentences were presented, cued and accompanied by one of three portraits that were labelled as the speakers' faces. In Block 1 speaker-specific probabilities of syntax errors were 10, 50, or 90%; in Block 2 they were uniformly 50%. In both blocks, speech errors elicited P600 effects in the scalp recorded ERP. We found a speaker sequence effect only in Block 1: the P600 to target words was larger after speaker switches than after speaker repetitions, independent of sentence correctness. In Block 1, listeners showed higher accuracy in judging sentence correctness spoken by speakers with lower error proportions. No speaker-specific differences in target word P600 and accuracy were found in Block 2. When speakers differ in error proneness, listeners seem to flexibly adapt their speech processing for the upcoming sentence through attention reorientation and resource reallocation if the speaker is about to change, and through proactive maintenance of neural resources if the speaker remains the same.
Keywords: adaptation; ERP; P600; sentence processing; sequential effect; speaker characteristics

1 Introduction

Speech processing requires adaptations to the speaker's identity. Prior EEG studies on spoken sentence processing have demonstrated that neural correlates of speech processing are modulated by stereotype-dependent inferences about the speaker, individuated by voice (van Berkum, van den Brink, Tesink, Kos, & Hagoort, 2008) or by accent (Goslin, Duffy, & Floccia, 2012; Grey & van Hell, 2017; Hanulíková, van Alphen, van Goch, & Weber, 2012; Romero-Rivas, Martin, & Costa, 2015; Xu, Abdel Rahman, & Sommer, 2019). More recent studies using faces to pre-cue speaker identity before each sentence found that seeing the speaker's face seems to help shaping linguistic expectations associated with the speaker identity in a proactive way (Grey, Cosgrove, & van Hell, 2020; Hernández-Gutiérrez et al., 2021). In studies using sentences produced by speakers with differential individual-specific communicative styles, newly-learnt speaker-specific characteristics in language use were found to shape listeners' syntactic expectations about particular sentence structures (e.g., Kroczek & Gunter, 2017 using auditory sentence presentation) or to elicit differential neural responses (e.g., P600 effects in ERPs after ironic expressions in a reading paradigm in Regel, Coulson, & Gunter, 2010). These findings suggest the existence of speaker-specific language representations that allow the listener to generate particular linguistic expectations and to adapt speech processing strategies accordingly (Kroczek & Gunter, 2020; see also discussion in Brown-Schmidt, Yoon, & Ryskin, 2015; Kuperberg & Jaeger, 2016). Therefore, listeners seem to learn and dynamically consider speaker-specific characteristics in language use when perceiving sentences produced by multiple speakers.

In experimental sessions, where participants are presented with spoken sentences produced by multiple unknown speakers, every perceived sentence can be viewed as a training stimulus that shapes the listener's knowledge about speaker-specific language use (MacDonald, 2013), for example, the probability of using particular sentence structures or committing speech errors. Since speaker-specific characteristics in language use seem to be learnt and considered during speech processing, listeners might need to match each perceived sentence to prior speaker-specific expectations by referencing the cumulated experience of the specific speaker in order to build and dynamically update their expectations about the language use of individual speakers. If that's the case, listening to multiple speakers would require calibration every time a speaker switch occurs, requiring attention control and recruiting memory resources (e.g., Kapadia & Perrachione, 2020; Wong, Nusbaum, & Small, 2004). However, the effect of speaker sequence (switch or repetition) across sentences while learning and adapting to local speaker-specific characteristics in language use on neural responses during spoken sentence processing has not been examined.

In fact, many studies on talker normalization indicated that increased attention control and working memory resources are required in multi-speaker as compared to single-speaker situations. In speeded word/vowel identification or judgment tasks, recognizing spoken words or vowels was slower and less accurate in multi-speaker as compared to single-speaker situations (Kapadia & Perrachione, 2020; Magnuson & Nusbaum, 2007; Nusbaum & Morin, 1992), and was accompanied by increased neurophysiological responses (Chandrasekaran, Chan, & Wong, 2011; Kaganovich, Francis, & Melara, 2006; Wong et al., 2004). Additionally, working memory for speech processing was shown to be susceptible to interference from variability across speakers in multi-speaker situations during a digit sequence recall task (Lim, Shinn-Cunningham, & Perrachione, 2019). Furthermore, in fMRI Wong and colleagues (2004) showed that word recognition in multi-speaker situations recruited not only classic speech areas (e.g., posterior superior temporal cortex) but also areas associated with attention shifts (superior parietal cortex). More recent studies demonstrated faster word identification in the current trial when the speaker was the same rather than a different person as in the preceding trial (Carter, Lim, & Perrachione, 2019; Kapadia & Perrachione, 2020; Lim, Tin, Qu, & Perrachione, 2019). Together, the findings about talker normalization in multi-speaker situations indicate that switching between multiple speakers triggers attentional reorientation and additional load on working memory (Lim, Shinn-Cunningham, & Perrachione, 2019; Shinn-Cunningham, 2008). Hence when perceiving sentences spoken by multiple speakers, neural correlates of speech processing may possibly show effects of speaker sequence (i.e., switch and repetition), which have yet to be investigated.

Effects of speaker sequence across sentences as well as effects of learning local speaker characteristics in language use might be reflected by variations in the P600 component in ERPs. The

P600 is usually maximal at centro-parietal scalp sites, starts around 500 ms after word onset and may extend to one second or more. P600 effects, that is, increased P600 amplitudes, are typically engendered by grammaticality violations (Friederici, Pfeifer, & Hahne, 1993), sentence ambiguities or high degrees of sentence complexity (Kaan & Swaab, 2003; Osterhout & Holcomb, 1992), but also by semantic violations, for example, semantic reversal anomalies (van Herten, Kolk, & Chwilla, 2005), or semantic blends (Xu et al., 2019). P600 effects in sentence processing are known to be susceptible to error probability, being attenuated under higher error probability, which may be stereotypically associated with a particular speaker group, for example, non-native as compared to native speakers, or produced by experimental manipulations (Coulson, King, & Kutas, 1998; Grey & van Hell, 2017; Hahne & Friederici, 1999; Hanulíková et al., 2012; Xu et al., 2019). Together, the consistency of P600 effects across various kinds of speech errors and ambiguities and its susceptibility to probability of speech errors indicate close interactions between language processing and cognitive control processes, for example, monitoring, attention, and working memory (Coulson et al., 1998; Hahne & Friederici, 1999; Kolk & Chwilla, 2007).

Various accounts have been suggested to explain the mechanisms behind P600 effects. Initially the P600 was related to syntactic reanalysis or repair (e.g., Friederici et al., 1993), but later it was suggested to reflect general reanalysis in speech perception (Kolk & Chwilla, 2007; Münte, Heinze, Matzke, Wieringa, & Johannes, 1998), or integration processes in sentence processing (Brouwer, Crocker, Venhuizen, & Hoeks, 2017). As a domain-general interpretation, the monitoring theory of language perception, explains the P600 in terms of conflict monitoring, an important aspect of cognitive control (van de Meerendonk, Indefrey, Chwilla, & Kolk, 2011; van de Meerendonk, Kolk, Chwilla, & Vissers, 2009; Vissers, Kolk, van de Meerendonk, & Chwilla, 2008). Monitoring is an evaluative element in cognitive control that keeps contextual information activated and entails the detection of conflicts and the triggering of compensatory adjustments (Braver, 2012; Botvinick et al., 2001). Closely related to monitoring, the P600-as-P3 account views the P600 as a variant of the P3b component (Coulson, 1998; Coulson et al., 1998). The P3b is elicited by uncertain, unexpected or surprising stimuli, for example, in the "oddball" and the task-switching paradigms (Jost, Mayr, & Rösler, 2008; see Leckey & Federmeier, 2020 for a review), and reflects the saliency of the stimuli

(Clayson & Larson, 2011; Donchin, 1981; Gratton et al., 2018; Johnson, 1986). The stimuli and contexts that are known to elicit and/or affect the P600 and P3b components are highly similar, for example, task demands, saliency of the stimuli, participant's attention, global and local subjective probability of the stimuli (Coulson, 1998; Coulson et al., 1998; Sassenhagen & Fiebach, 2019; Sassenhagen, Schlesewsky, & Bornkessel-Schlesewsky, 2014). Despite the ongoing debate about the cognitive functions of the P600, the overview of prior studies suggests that the P600 is a robust marker for understanding the online processing of language processes and how that processing changes with experience and context, which possibly engages cognitive control mechanisms.

As mentioned above, P600 effects are sensitive to stereotype-driven speaker characteristics indicated by voices and accents (Grey & van Hell, 2017; Hanulíková et al., 2012; Romero-Rivas et al., 2015; van Berkum et al., 2008; Xu et al., 2019). Concerning short-term speaker-specific adaptations, a previous EEG study using a reading paradigm conducted by Regel, Coulson, and Gunter (2010) assessed how newly-learnt speaker-specific communicative styles affect sentence processing. They manipulated the proportion of ironic statements in short passages of written text produced by two individuals in two sessions and found differential P600 effects after ironic statements made by the two individuals (Regel et al., 2010). Therefore, the P600 might be a reliable indicator of short-term experience with speaker-specific characteristics during language processing. Furthermore, the P600 is susceptible to error probability of a particular speaker group or experimental situation (Hahne & Friederici, 1999; Hanulíková et al., 2012; Xu et al., 2019). Thus, in the present study, we wanted to investigate how speaker-specific error probability may generate differential speaker-specific P600 effects.

To conclude, prior studies showed that the P600 is linked to cognitive control in language processing (e.g., Xu, Abdel Rahman, & Sommer, 2021), and that the P600 is sensitive to stereotype-driven speaker characteristics and error statistics in local environments (e.g., Xu et al., 2019) and to speaker-specific communicative styles (Regel et al., 2010). Therefore, P600 variations can be invoked to assess speaker-specific processing strategies based on newly-learnt individual-specific characteristics in error probability and to investigate effects triggered by speaker sequences in multi-speaker situations that possibly engage cognitive control mechanisms.

1.1 The Present Study

Complementing previous research, the present study aimed to investigate two aspects of adaptations to speaker identity during speech processing in multi-speaker situations – the effects of speaker sequence across sentences and of learning local speaker-specific error probabilities. In each trial, participants were presented with a photographic portrait before and during listening to a sentence spoken in the voice attributed to the person on the picture; the sentences could be correct or contain a syntactic error. Each experimental session was divided into two blocks. In Block 1, three native speakers (identified by face and voice pairs) were associated with different proportions of sentences with errors (10%, 50%, and 90%); in Block 2, the same three speakers were employed but now all committed 50% errors. Randomly interspersed in 1/3 of all trials, a correctness judgment of the immediately preceding sentence was required.

Notably, complementing the voice as speaker identity-defining property, we used three photographic portraits as additional cues preceding and accompanying the sentences. In previous studies about talker normalization and in most EEG studies about spoken sentence processing, speaker identities were only indicated by differential acoustic-phonetic properties without any visual cues such as photographic portraits of the speakers. Face-to-face communication is natural in daily life, and listeners normally can tell who is going to speak next based on facial expression or body language. Recent studies have shown that associating and cueing speaker identity with a portrait may modulate P600 effects after syntactic violations (Grey et al., 2020) or N400 effects after semantic violations (Hernández-Gutiérrez et al., 2021). Portraits before sentence presentation seem to help shaping linguistic expectations associated with the speaker identity in a proactive way. If listeners indeed calibrate their linguistic expectations and processing strategies through attention control for a switched speaker and adapt speech processing strategies for an upcoming sentence according to newly-learnt speaker-specific characteristics, pre-cueing each sentence using a photographic portrait should enhance the hypothesized effects of speaker sequence and speaker-specific error probability. The specific research questions are described in the following.

1.1.1 Speaker Sequence Effect

Our first research question was whether speaker sequence (switch or repetition) across consecutive sentences would affect the processing of upcoming sentence(s) or not. In a previous EEG study about spoken sentence processing, a sequential effect of previous-sentence error detection on the current-sentence P600 was found: after detecting a speech error in the preceding sentence, the P600 amplitude in the current sentence decreased for both correct and incorrect critical words (Xu et al., 2021). This sequential effect was taken to reflect heightened attention and sustained proactive control over the monitoring of upcoming sentences, triggered by detecting a speech error in the preceding sentence. If listeners indeed calibrate their linguistic expectations and processing strategies through attention control for a switch in the speaker identity across sentence borders, similar sequential effects in P600 amplitude should be observed in the present study.

We hypothesized that if the sentence-preceding face cues indicate a switch in speaker identity for the upcoming sentence, listeners would reorient their attention to the new speaker and activate additional resources in working memory. Meanwhile, cued speaker repetitions might “prime” previously activated resources for the person who continues speaking, that is, neural resources might be recruited and maintained in advance for processing the next sentence – similar to the sequential effects reported by Xu et al. (2021). To sum up, after a speaker switch we expected larger P600 amplitudes after critical words in the current sentence irrespective of its correctness, as compared to a speaker repetition across consecutive sentences.

1.1.2 Speaker Identity Effect

Our second research question was whether short-term experience with individual speakers associated with different error probabilities would persistently shape speech processing strategies for these speakers. By manipulating speaker-specific error proportions in Block 1 of the present experiment, participants could cumulatively collect experience with speaker-specific error probability. Participants should associate higher error rates of a given speaker with diminished linguistic competence of that speaker. If the newly-learnt speaker-specific error probability enduringly shifts speech processing strategies for each individual speaker as hypothesized, differential P600 effects should be observed for the three speakers in Block 2, although they would now show the same error proneness. Thus, we expected to see smaller P600 effects in Block 2 for the speaker(s) that had been deemed less competent, i.e., had shown higher error proportions in Block 1.

In addition to P600 effects, accuracy of sentence correctness judgments might be affected by speaker-specific error probability in Block 1. Thus, we expected higher accuracy in sentences correctness judgments for the more competent as compared to the less competent speaker(s) in Block 1 and possibly also in Block 2.

1.1.3 Face Processing

In order to provide complementary evidence that the speaker's identity associated with each face is learnt and recognized by the participants, the ERP components N170 and N250 elicited by the face cues were analysed. The N170 is a large bilateral temporo-occipital negativity, peaking between 150 and 190 ms after stimulus onsets, generally taken to reflect the structural encoding of faces, because it is usually larger for faces than other visual objects (Bentin, Allison, Puce, Perez, & McCarthy, 1996; Eimer, 1998). The N170 is systematically suppressed (or adapted) by the repetition of faces as a stimulus category, i.e., when a face is preceded by another face, irrespective of its identity (for reviews see Rossion & Jacques, 2011, Rossion, 2014). But the N170 is also found to be suppressed by the repetition of individual face identities, as compared to face identity switches (e.g., Jacques & Rossion, 2006). Thus, the N170 is suggested to reflect face-category detection as well as adaptation to individual face identities. The N250 is an ERP component with an occipito-temporal distribution, starting between 180 and 220 ms and peaking between 230 and 330 ms after face onsets (Schweinberger, Pfütze & Sommer, 1995; for review see Schweinberger & Neumann, 2016). The N250 is larger to repeated (i.e., primed) as compared to non-repeated or novel (i.e., unprimed) faces across successive presentations, being more prominent over the right hemisphere; this priming effect is known as the N250r and is larger for familiar than unfamiliar faces (Dörr, Herzmann, & Sommer, 2011; Herzmann, Schweinberger, Sommer, & Jentzsch, 2004; Schweinberger et al., 1995).

In the present study, the faces were shown already before the sentence and terminated together with the sentence. If speakers' identities associated with the faces were learnt during the experiment, a repetition of a face identity between consecutive trials as compared to a switch into a different face should diminish the N170 – reflecting detection and sequential adaptation to face identities (e.g., Jacques & Rossion, 2006), but might enhance the N250, i.e., engender an N250r effect – reflecting repetition priming of individual face identities (e.g., Dörr et al., 2011; Herzmann et al., 2004).

2 Method

2.1 Participants

Thirty-nine native German speakers (27 women, mean age = 26 years, *SD* = 4) without hearing, neurological, or psychiatric disorders were tested at the psychology department in Humboldt-Universität zu Berlin. All participants had normal or corrected-to-normal visual acuity and normal colour vision according to self-report and were right-handed according to the Edinburgh Questionnaire (Oldfield, 1971). All participants gave informed consent and received payment or course credit for participation.

2.2 Materials

Materials consisted of 180 sentences with grammatical agreement violations and their well-formed counterparts. All sentences were spoken by a female German speaker pronouncing in standard German with neutral intonation at normal speed. A total of 360 audio files were recorded in a professional studio using a Neumann® TLM 103 condenser microphone with fixed heart-shaped directivity. Sentences were digitized with 44.1 kHz at 24 bit resolution and stored in wave-format. GoldWave® v5.70 software was used to change the pitch of the speaker into three different voices that sounded natural and distinguishable from each other, and to mark the onsets of critical events in each sound file. Mean sentence duration was 3.3 s (*SD* = 0.75) and did not vary across speaker conditions. No critical word in any sentence was at the first or last word position.

Portraits of three Caucasian faces taken from the FACES database (Ebner, Riediger, & Lindenberger, 2010; Lindenberger, Ebner, & Riediger, 2005-2007) were used as visual cues indicating the speaker’s identity. All portraits were in grey scale, cut into a square format with only the face filling the square, showed neutral expressions and direct gaze towards the viewer.

Throughout the experiment each voice was consistently assigned to one of the faces, resulting in three face-voice pairs. Magnuson and Nusbaum (2007) have shown that when expecting multiple speakers in a given situation, even without any visual cues, listeners still showed adaptation effects to the variability across speakers that was conveyed only by slight differences in voice pitch. Therefore, the artificially manipulated voices with different pitches that coupled with differential photographic portraits respectively should be adequate for the purpose of differentiating the three speaker identities.

Notably, speech errors consisted of different types of grammatical agreement violations: gender agreement violations between determiner and noun, number agreement violations between subject/pronoun and verb or between determiner and noun, and case agreement violations between verb and object (for details see Xu et al., 2019, Exp. 2). The high variability in error type, position and sentence structure aimed to counteract subjective control or adaptation based merely on superficial sentence or error structures. Examples of each type of sentence are provided in Table 1.

TABLE 1

2.3 Procedure

The experimental session consisted of two blocks, each allowing for nine in-between breaks of participant-determined duration. All 180 sentences were presented in Block 1 and repeated in Block 2. Within a given block there was no repetition of any sentence (correct or incorrect) or its counterpart. There were three speakers (face-voice pair), and each speaker uttered 60 sentences per block. In the first block, the 10%-speaker spoke 10% ($N = 6$) of all sentences incorrectly, the 50%-speaker spoke half of the sentences ($N = 30$) incorrectly, and the 90%-speaker spoke 90% of all sentences ($N = 54$) incorrectly. In the second block, every speaker committed 50% speech errors. The overall error proportion in each block was thus 50%. The association between speaker (face-voice pair) and error proportion in Block 1 was counterbalanced over participants. Within each block, sentences were randomly assigned to one of the three speakers and to be correct or incorrect according to the corresponding error proportion of each speaker in a given block. That means, independently for Block

1 and Block 2, errors were randomly selected from the sentence pool, thus incorrect sentences in Block 1 could be correct or incorrect in Block 2 and vice versa. Sentences in Block 2 were presented in a differently randomized order than in Block 1. Speakers were randomized in different orders, with the constraint that the conditions of speaker switch and speaker repetition across consecutive trials occurred equally often in each block.

Participants were tested in a sound-attenuated chamber. Visual stimuli were presented on a 19-inch DELL® 1908 FPb monitor; the sentences were presented using two Creative® Gigaworks T20 speakers placed at both sides of the monitor. Audio volume was adjusted to a clear and comfortable level for each participant before the experiment.

Each trial began with a picture of a face, cueing speaker identity. After 800 ms, the spoken sentence was presented while the face picture remained on the screen. The picture terminated together with the sentence. Participants were instructed not to blink while the face was on the screen, fixate the face and listen to the sentence for understanding. They were told that the face showed the speaker of the upcoming sentence. In 1/3 of all trials, that is, in 60 trials per block and randomly interspersed, a correctness judgment about the immediately preceding sentence was required by pressing one of two buttons placed on the desk within 3 s. Of 39 participants, 20 pressed the left button for “correct” judgments, and 19 pressed the right button.

After the experiment, a short calibration procedure recorded prototypical eye movement artefacts, to be used for later correction. Following the task, participants were orally debriefed about whether they had noticed a difference in the error proportions associated with the different speakers in the two blocks, and whether the voices had sounded natural and distinguishable.

2.4 Electrophysiological Recordings

The EEG was recorded from 64 Ag/AgCl electrodes arranged according to the extended 10/20 system. The left mastoid served as initial reference. We used electrodes near the left and right canthi of both eyes and above and beneath the left eye to register eye movements and blinks. Impedances of all electrodes were kept below 5 kΩ. The EEG signals were filtered online at a band pass of 0.1-1000 Hz. Offline, the EEG was re-calculated to average reference and low-pass filtered at 30 Hz (24 dB/oct). Eye movement and blink artefacts were corrected employing BESA® software (Berg & Scherg, 1994). The continuous signals were segmented into epochs of 1.3 s, starting 100 ms before the onset of the critical events in the target sentences and after the cue faces; these 100 ms served as baseline. EEG segments with voltage ranges exceeding 100 µV were excluded. For grammatical agreement violations and their corresponding correct versions, markers for EEG segmentation were placed at the onsets of critical words, at which the ungrammaticality of the sentence became apparent, if present. For cue-locked analyses, ERPs were synchronized to onsets of the cue faces. Voices should not have affected the face processing ERPs, because face presentation started 800 ms before the onset of the audio signal, by which time the face-elicited N170 and N250 would have subsided. Finally, EEG segments were averaged separately for each condition, block, electrode, and participant. All EEG processing steps were conducted using MATLAB® R2016a software and the toolboxes EEGLAB (Delorme & Makeig, 2004) and FieldTrip (Oostenveld, Fries, Maris, & Schoffelen, 2011) in a 64-Bit Windows® 7 operating system.

2.5 Data Analysis

The P600 was analysed in a centro-parietal ROI of 25 electrodes (C1/2, C3/4, CP1/2, CP3/4, CP5/6, P3/P4, P5/6, PO3/4, PO7/8, O1/2, Cz, CPz, Pz, POz, Oz) during a 600-1200 ms time window, as in Xu et al. (2019, 2021)[1]. Mean amplitudes in the P600 ROI averaged across all conditions and trials were calculated for each participant. Upon closer inspection of the data, one data set was extremely different from the others because the P600 amplitude of this data set was 1.98 µV and 3.9 *SD* above the average (N = 39, M = 0.35 µV, SD = 0.42 µV). Therefore, this data set was considered as outlier and excluded from further analyses[2] (Weissgerber, Milic, Winham, & Garovic, 2015).

All analysis of variance (ANOVA) used repeated measures factors. For all *post hoc* analyses, Bonferroni correction was employed.

2.5.1 Speaker Sequence Effect

To test effects of speaker sequence on current-sentence processing and to directly compare speaker-sequence effects between the two blocks, we conducted an ANOVA with factors block (Block 1 vs. 2), speaker sequence (switch vs. repetition between consecutive sentences), and sentence correctness (correct vs. incorrect) on the mean P600-ROI amplitudes.

2.5.2 Speaker Identity Effect

In order to test the transfer of speaker-specific experience gained in Block 1 to speech processing in Block 2, an ANOVA with factors speaker identity (the 10%-, 50%-, 90%-speaker) and sentence correctness (correct vs. incorrect) was performed on the mean P600-ROI amplitudes for Block 2 ERPs. Please note that the factor speaker identity in all analyses reported here represents the speaker identity that had been associated with either 10%, 50%, or 90% error proportions in Block 1, not the face-voice pairs per se.

In addition, accuracies in the correctness judgment task were calculated for each speaker identity in each block and entered into an ANOVA with factors block (Block 1 vs. 2) and speaker identity (the 10%-, 50%-, 90%-speaker). This analysis was conducted to provide evidence that the error statistics associated with each speaker were indeed learnt and affected the accuracy of sentence correctness judgments in each block.

2.5.3 Face Processing

Mean N170 and N250 amplitudes were calculated in the time windows 160-230 and 280-350 ms after cue onset, respectively, in occipito-temporal ROIs of three electrodes at left and right hemispheres (P5, P7, TP9 vs. P6, P8, TP10). The N170/N250 ROI and time window were based on the literature (e.g., Herzmann et al., 2004; Jacques, d'Arripe, & Rossion, 2007; Schweinberger, Pickering, Jentzsch, Burton, & Kaufmann, 2002; Wiese et al., 2019). Visual inspection confirmed that the time windows were reasonable for detecting both effects, that is, avoiding temporal overlap between the two ERP components. Within the N170 and N250 ROIs amplitudes were averaged across electrodes and submitted to ANOVAs with factors block (Block 1 vs. 2), hemisphere (left vs. right hemisphere), and speaker sequence (switch vs. repetition between consecutive sentences).

2.5.4 Additional Analysis

Because in the current design, only the 50%-speaker condition had the same trial numbers in correct and incorrect sentence conditions in each given block. In order to assess possible differences between blocks due to participants' fatigue or practice, a further ANOVA was conducted with factors block (Block 1 vs. 2) and sentence correctness (correct vs. incorrect) on P600 amplitudes of the 50%-speaker in both blocks. Besides, in order to reveal any fatigue or practice effects across blocks as well as to address concerns about any artefacts caused by the face-voice pairs used, the accuracy of correctness judgments was calculated for each face-voice pair in each block and entered into an ANOVA with factors block (Block 1 vs. 2) and face-voice pair (Face-voice 1, 2, and 3). Please note that the factor face-voice pair does not correspond to the factor speaker identity, that is, the three speaker identities that had been associated with different error proportions in Block 1.

3 Results

Mean accuracy in the correctness judgment task was 89.13 % (*SD* = 8.45 %). In the oral debriefing after the experiment, all participants confirmed that the voices had sounded natural and distinguishable, and that they had noticed the differences in the error proportions of the speakers and between blocks.

3.1 Speaker Sequence Effect

Figure 1 shows ERP waveforms and mean amplitudes from the P600 ROI in the various conditions. The ANOVA confirmed larger P600 amplitudes to incorrect than to correct words (M = 0.41 µV vs. 0.20 µV; $F(1, 37) = 11.49$, $p = .002$, $\eta_p^2 = .237$). There was also a significant main effect of block ($F(1, 37) = 9.98$, $p = .003$, $\eta_p^2 = .212$): P600 amplitudes in Block 2 were larger than in Block 1 (M = 0.38 µV vs. 0.22 µV). Most importantly, the factor speaker sequence interacted with block ($F(1, 37)$ = 5.34, $p = .027$, $\eta_p^2 = .126$). Pairwise comparisons revealed that speaker sequence was only significant in Block 1 (Block 1: $F(1, 37) = 7.78$, $p = .008$, $\eta_p^2 = .174$; Block 2: $F(1, 37) = .002$, $p = .964$): P600 amplitudes to both correct and incorrect words on Trial N were larger when the speaker had changed from Trial N-1 to the current Trial N than when the speaker was the same (M = 0.33 µV vs. 0.12 µV). No other effects were found (see Table 2 for further details).

FIGURE 1

TABLE 2

3.2 Speaker Identity Effect

In the ANOVA of speaker identity-specific P600 effects in Block 2, the factor sentence correctness was significant ($F(1, 37) = 10.61$, $p = .002$, $\eta_p^2 = .223$), confirming that incorrect words engendered larger P600 amplitudes than correct words ($M = 0.55$ µV vs. 0.25 µV). No other effects were found (see Table 3 for further details).

TABLE 3

The ANOVA of accuracy with factors block and speaker identity revealed a significant main effect of speaker identity ($F(2, 74) = 6.60$, $p = .002$, $\eta_p^2 = .151$), indicating that the accuracy of sentence correctness judgments decreased from the 10%- over the 50%- to the 90%-speaker. Mean accuracy and standard deviation in each condition are displayed in Table 4. The ANOVA also revealed a significant interaction between speaker identity and block ($F(2, 74) = 8.06$, $p = .001$, $\eta_p^2 = .179$). Pairwise comparisons showed significant differences between accuracies for the three speakers in Block 1 ($F(2, 36) = 10.92$, $p < .001$, $\eta_p^2 = .378$), but not in Block 2 ($F(2, 36) = 1.25$, $p = .298$, $\eta_p^2 = .065$). In Block 1, sentence judgment accuracy was higher for speakers with lower error proportions (Table 4). Factor Block did not yield a main effect ($F(1, 37) = .48$, $p = .492$, $\eta_p^2 = .013$).

TABLE 4

FIGURE 2

Mean P600 amplitudes averaged separately for each participant in each speaker identity and sentence correctness condition in Block 1 and 2 are displayed in Figure 2B. In Block 1, errors spoken by the 10%-speaker seemed to engender larger differences between P600 amplitudes generated by incorrect and correct words compared to 50%-speaker, whereas the 90%-speaker seemed to show a reversed effect in Block 1 – increased P600 amplitudes after correct words compared to errors. In

Block 2, all three speakers engendered similar P600 effects after errors compared to correct words, which was confirmed by the previous ANOVA. Because in Block 1 there were at most 6 trials each for incorrect sentences in the 10%-speaker and for correct sentences in the 90%-speaker, a statistical comparison between Block-1 and Block-2 P600 effects with the factor speaker identity (10%-, 50%-, and 90%-speaker) was not planned at first. However, results from the accuracy analysis indicated that participants might have used different processing strategies for the speakers in Block 1. Hence, we decided to conduct an exploratory analysis of P600 amplitudes in Block 1 despite the partially unequal trial numbers. An ANOVA with factors speaker identity (the 10%-, 50%-, 90%-speaker) and sentence correctness (correct vs. incorrect) was performed on the mean P600-ROI amplitudes for Block 1; only a trend was found for the interaction between sentence correctness and speaker identity ($F(2, 74) = 2.54$, $p = .086$, $\eta_p^2 = .064$). No other effects were found (see Table 5 for further details). Although the interaction was not significant, which might be due to the extremely small trial numbers in some conditions, we conducted pairwise comparisons to explore any speaker identity effect in P600 effects in Block 1, and found that the factor sentence correctness was significant for the 10%-speaker ($F(1, 37) = 4.82$, $p = .034$, $\eta_p^2 = .115$) but not for the other two speakers in Block 1[3] (see Table 5 for further details).

TABLE 5

3.3 Face Processing

As shown in Figure 3, there was a slow positive wave superimposed on the face-elicited ERP, shifting the amplitudes of the negative-going N170 and the N250 components into the positive range. For N170 amplitudes, speaker sequence showed a strong main effect ($F(1, 37) = 19.76$, $p < .001$, $\eta_p^2 = .348$); the amplitude in the occipito-temporal ROI during the 160 to 230 ms time segment was less negative for face repetitions than face switches ($M = 2.30$ µV, $SD = 2.12$ µV vs. $M = 1.88$ µV, $SD = 2.18$ µV), reflecting a face identity adaptation effect of the N170 (Fig. 3). No other effects were found (see Table 6 for further details).

In N250 amplitudes, speaker sequence had a strong main effect ($F(1, 37) = 60.35$, $p < .001$, $\eta_p^2 = .620$); the amplitude in the occipito-temporal ROI during the 280 to 350 ms time segment was more negative for face repetitions than switches ($M = 2.96$ µV, $SD = 2.02$ µV vs. $M = 3.95$ µV, $SD = 2.28$ µV), reflecting the typical priming effect, i.e., the N250r. Hemisphere showed a main effect in both blocks ($F(1, 37) = 12.55$, $p = .001$, $\eta_p^2 = .253$), being larger in the right than the left hemisphere ($M = 3.91$ µV, $SD = 2.47$ µV vs. $M = 3.00$ µV, $SD = 1.84$ µV). There was also an interaction of block and speaker sequence ($F(1, 37) = 11.05$, $p = .002$, $\eta_p^2 = .230$); the priming effect, i.e., the difference between speaker switch and repetition, seemed to be larger in Block 1 than Block 2 (Block 1: $M_{repetition} = 2.93$ µV, $SD = 2.04$ µV vs. $M_{switch} = 4.08$ µV, $SD = 2.33$ µV; Block 2: $M_{repetition} = 2.99$ µV, $SD = 2.00$ µV vs. $M_{switch} = 3.82$ µV, $SD = 2.23$ µV). Pairwise comparisons on the factor speaker sequence showed that this effect was significant in both blocks (Block 1: $F(1, 37) = 64.23$, $p < .001$, $\eta_p^2 = .634$; Block 2: $F(1, 37) = 41.26$, $p < .001$, $\eta_p^2 = .527$) (Fig. 3); pairwise comparisons on the factor block showed that block was not significant for either speaker sequence (Table 6). No other effects were found (see Table 6 for further details).

FIGURE 3

TABLE 6

3.4 Additional Analysis

The ANOVA comparing the 50%-speaker conditions between blocks revealed a main effect of block ($F(1, 37) = 5.37$, $p = .026$, $\eta_p^2 = .127$), which also showed that P600 amplitudes for the 50%-speaker were larger in Block 2 ($M = 0.42$ µV) than in Block 1 ($M = 0.19$ µV). Although incorrect words engendered larger mean P600 amplitudes than correct words spoken by the 50%-speaker overall ($M = 0.38$ µV vs. 0.23 µV), this effect was not significant ($F(1, 37) = 2.88$, $p = .098$, $\eta_p^2 = .072$). No interaction between block and correctness was found for this speaker ($F(1, 37) = .92$, $p = .344$, $\eta_p^2 = .024$). The nonsignificant main effect of sentence correctness might be due to relatively small P600 effects in Block 1; the difference between incorrect and correct sentence conditions of the 50%-

speaker in Block 1 seemed to be very small and variable ($M = 0.22$, $SD = 0.63$ μV vs. $M = 0.16$, $SD = 0.54$ μV) compared to Block 2 ($M = 0.55$, $SD = 0.69$ μV vs. $M = 0.29$, $SD = 0.65$ μV) (Fig. 2), which might have rendered the overall P600 effect for this speaker into a trend.

The ANOVA on accuracy with factors face-voice pair and block showed that on average accuracy did not vary across face-voice pairs ($F(1, 37) = .37$, $p = .696$, $\eta_p^2 = .010$), being $M = 89.75\%$ ($SD = 8.51\%$), 88.96% (8.43%), and 88.68% (8.56%) for each face-voice pair, respectively, indicating that isolated from the manipulation of error statistics, the face and voice stimuli per se did not have a significant influence on sentence processing. Overall accuracy also did not vary across blocks ($F(1, 37) = .48$, $p = .492$, $\eta_p^2 = .013$), being 88.74% ($SD = 8.57\%$) in average in Block 1 and 89.52% ($SD = 8.43\%$) in average in Block 2, which argues against fatigue or practice effects. Nor was there an interaction between block and face-voice pair ($F(2, 74) = .32$, $p = .731$, $\eta_p^2 = .008$).

4 Discussion

The current study addressed two aspects of adaptation to speaker identity during processing spoken sentences in multi-speaker situations: the effect of speaker sequences across sentences and learning speaker-specific error probabilities. Overall ERP results confirmed the typical P600 effect to speech errors relative to correct words. In Block 1, where speakers were associated with individualized error statistics, P600 amplitudes after critical words in both correct and incorrect sentences were smaller when the current speaker was the same person as in the previous sentence as compared to when the speaker had changed. Also, the accuracy of sentence correctness judgments was inversely related to speaker-specific error probability in Block 1. However, in Block 2, where error probabilities of the speakers were equal, no speaker sequence effect on the P600 and no effect of previously-learnt speaker-specific error probability on the P600 effect and on the accuracy of judgments were found.

4.1 Speaker Sequence Effect

In Block 1 we found a sequential adaptation effect in the P600 triggered by the speaker sequence across sentence borders. As explained in the introduction, when perceiving sentences produced by initially unfamiliar speakers in a multi-speaker situation, in order to economically allocate resources and optimally adapt to the speakers, listeners may use the speaker identity (face, voice, etc.) as a contextual cue and every perceived sentence as a training stimulus to form and update their speaker-specific expectations in a cumulative manner (Brown-Schmidt et al., 2015; Kleinschmidt & Jaeger, 2015; Kuperberg & Jaeger, 2016). From this perspective, on the one hand, a switch in speaker identity can be considered as a switch in the context for the next sentence, disrupting the attentional focus to a given speaker/context and imposing a load on working memory; listeners may need to refer to speaker-specific expectations and reallocate neural resources for a switched speaker, hence increasing the P600 amplitudes in the upcoming sentence. This may especially be the case when the identity of the impending, turn-taking speaker is pre-cued, for example, by a visual signal, such as the portraits in the present design.

On the other hand, the present adaptation effects triggered by speaker sequence may also be attributed to the repetition of speaker identity across sentences. Neural resources for speech processing may be proactively maintained (or “primed”) when knowing from the face cue that the same speaker will continue to speak (see Xu et al., 2021 for similar proactive adaptation in spoken sentence processing reflected in sequential effects in the P600). One of the most robust experience-related cortical dynamics induced by stimulus repetition is repetition suppression, that is, the reduction in neural activity, commonly linked to performance improvements due to repetition or priming (for a review see Grill-Spector, Henson, & Martin, 2006). Although the neural causes underlying repetition priming or suppression are still debated, several models suggest similar maintenance of neural activities across stimulus repetitions explaining these effects. For example, the Fatigue model suggests diminished overall activation for stimulus repetition (Grill-Spector & Malach, 2001; Miller & Desimone, 1994), and the Sharpening model suggests fewer neurons responding after stimulus repetition (Desimone, 1996; Li, Miller, & Desimone, 1993; Wiggs & Martin, 1998). In any case, listeners may use the speaker identity as a contextual cue and proactively maintain the neural resources for a repeated speaker of the upcoming sentence, for example, the neural resources for speech monitoring, reflected in P600 amplitudes.

The proposed interpretation of speaker sequence effects is in line with a feedforward auditory streaming model of speaker adaptation, which was initially developed to explain findings about talker normalization in multi-speaker situations (Carter et al., 2019; Kapadia & Perrachione, 2020; Lim, Shinn-Cunningham, & Perrachione, 2019; Shinn-Cunningham, 2008). Accordingly, the switch between speakers across consecutive trials imposes attentional reorientation, whereas speaker repetition facilitates speech processing in a feedforward manner (Shinn-Cunningham, 2008; Lim, Shinn-Cunningham, & Perrachione, 2019). This model was developed based on findings in paradigms presenting spoken words or vowels rather than sentences, and without pre-cueing speaker identity; hence, this model was not intended to explain sentence processing mechanisms. Nevertheless, the idea that speaker identity (via face or voice) can be conceptualised as contextual cue and listeners dynamically adapt their expectations when listening to multiple speakers is very similar. Consistent with this model, we therefore suggest that the speaker sequence effect found in the present Block 1 should be attributed to resource allocation through attentional reorientation and load on working memory when the speaker switches across trials, as well as proactive maintenance of neural resources when the same speaker continues to speak.

The current findings seem to be in accord with accounts that relate the P600 to cognitive control mechanisms, for example the monitoring theory of language perception (van de Meerendonk et al., 2009) and the P600-as-P3 account (Coulson et al, 1998; Leckey & Federmeier, 2020). In the present results, the P600 showed similar patterns of trial-to-trial modulations triggered by speaker sequence as the P3b component in task-switching paradigms: P3b amplitudes were larger for task switch trials relative to repeat trials (Kopp, Steinke, & Visalli, 2020; see Kiesel et al., 2010 for a review). Task-switching paradigms and the various speaker identities in the present paradigm may both be viewed as contexts that require differential processing strategies to cope with in each trial/sentence. From this perspective, future studies may, for example, use non-face objects as cues to investigate whether speaker identities are necessary to induce particular linguistic expectations and trigger sequential adaptation in spoken sentence processing.

4.2 Speaker Identity Effect

Contrary to our initial predictions, adaptation to speaker characteristics reflected by speaker sequence effects and differential accuracy in judgments between speakers were only found in Block 1 and did not carry over to Block 2. Also, no speaker-specific P600 effects were found in Block 2 where speakers did not differ any more in their error probability. Notably, the overall invariant accuracy in sentence correctness judgments across blocks as well as increased P600 amplitudes in Block 2 relative to Block 1 (see Fig. 2; in line with results reported by Xu et al. (2019)) argue against any influence of fatigue over time. Hence it appears plausible that the absence of speaker sequence and speaker identity effects in Block 2 indeed relates to the contrasts between differences in the speakers' error proneness in the two blocks.

One possible interpretation is that listeners noticed (whether consciously or not) and quickly adapted to the now indistinguishable error statistics across the three speakers in Block 2. Novel environmental statistics have been reported to rapidly overwrite previous experience, abolishing or reversing previous effects (e.g., Regel et al., 2010). As discussed above, considering the speaker identity as contextual cues, listeners may dynamically estimate the reliability of their cumulatively-formed prior expectations and update their expectations accordingly. If the new input is inconsistent with prior beliefs, the representations and expectations concerning the speakers may then be adjusted. Hence, after noticing the now indistinguishable error probabilities in Block 2, listeners may have estimated the prior speaker-specific beliefs to be unreliable, thus not considering the speaker identity as reliable cues for adjusting speech processing strategies for sentences in Block 2 anymore, and (intentionally or not) employing the same or similar expectations and speech processing strategies for the speakers. Such strategic control could possibly abolish the expected transfer effects of speaker-specific error probability from Block 1.

Importantly, not observing transfer effects in Block 2 does not necessarily offer evidence against speaker-specific processing strategies in Block 1. In Block 1 the accuracy of the participants was found to be inversely related to the error probability associated with each speaker; the numeric patterns also suggested variability in P600 effects across speakers in Block 1 (Fig. 2; indicated by a trend in the exploratory analysis on the Block 1 P600). Hence speaker-specific processing strategies might have been adapted in Block 1. While nonsignificant, the enhanced P600 effect for the speaker with lower error probability in Block 1 coupled with equivalent P600 effects for all speakers in Block 2 are in keeping with the interpretation that speakers did use *locally-available* information about the speakers' reliability to adapt speech processing.

Nevertheless, the absence of speaker-sequence effects in Block 2 may indicate that the effects observed in Block 1 did not merely rely on switches versus repetitions on speaker identity per se but depended on differences in certain properties of the speakers (e.g., error proneness), which were associated with the identity-conveying information such as face and voice. Possibly, the more differential these properties are, the more likely the speaker identity is to be used as contextual cues and affects subsequent language perception. Previous studies that found transfer effects of speaker-specific language properties used only two speaker identities in an experimental block or session (spoken sentences: Kroczek & Gunter, 2017; reading paradigm: Regel et al., 2010). This may have enhanced the contrast in the differential proportions of a particular sentence structure between the two speakers, leading to transfer effects to a novel situation with neutral distribution. For subtle differences like error proportions between more than two speakers without differentiating error types or sentence structures as used in the current design, for transfer effects of speaker-specific processing to take place, there must be a meaningful difference between speakers that inherently affects processing in the *present* situation. If we included only two speakers for larger contrast in one experimental session, or manipulated the error types produced by each speaker, we might be able to see carry-over effects in the subsequent neutral block. Future studies may also try associating different error types (e.g., semantic replacement errors vs. syntactic errors) with different speakers using a similar paradigm as the present study; differential N400 and P600 effects may be expected for the different speakers.

4.3 Face Processing

Last but not least, face cue-locked analyses revealed non-linguistic effects of adaptation and priming to face identities. Above all, these face-specific effects of speaker identity and sequence indicate that speaker identities associated with the faces had been learnt, processed and adapted to in both blocks, supporting our claim that the speaker sequence effect in Block 1 was indeed influenced by repetition or switch of speaker identity between consecutive trials. The present study is an interesting case where at the same time, an amplitude-attenuating adaptation effect in the N170 and an amplitude-enhancing priming effect in the N250 have been observed. More specifically, N170 amplitude was diminished by the repetition of the same face (or picture). This is in line with the suggestion that N170 reflects multiple face processing stages including categorical but also individual face discrimination (Jacques & Rossion, 2006; Rossion & Jacques, 2011). We speculate that the adaptation effect was enabled by the long presentation time of the individual faces (3.35 to 4.85 s), starting in advance of the sentence and ending together with the sentence. This corresponds to a long adaptation period before the next face was presented. Face repetitions engendered larger N250 amplitudes than switches, reflecting the well-known repetition priming to face identities (Schweinberger et al., 1995; Schweinberger & Neumann, 2016).

5 Conclusion

In conclusion, the present study offers insights into when and how listeners adapt to unfamiliar speakers during processing spoken sentences in multi-speaker situations. Our findings suggest that listeners cumulatively learn speaker-specific characteristics in language use and dynamically adapt to speaker sequences and to the impending speaker in situations where the speakers differ in their language use. And cognitive processes of attention and memory are engaged when adapting to speaker identity across sentence borders during speech processing. Whether speaker identity is considered as reliable contextual cues for adapting speech processing strategies or not, however, seems to be determined by the degree of contrasts/differences in language use between the speakers in one local environment.

References

Barrett, S., & Rugg, M. D. (1989). Event-related potentials and the semantic matching of faces. *Neuropsychologia, 27*(7), 913-922. https://doi.org/10.1016/0028-3932(89)90067-5

Barrett, S. E., Rugg, M. D., & Perrett, D. I. (1988). Event-related potentials and the matching of familiar and unfamiliar faces. *Neuropsychologia, 26*(1), 105-117. https://doi.org/10.1016/0028-3932(88)90034-6

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience*, 8(6), 551-565. https://doi.org/10.1162/jocn.1996.8.6.551

Berg, P., & Scherg, M. (1994). A multiple source approach to the correction of eye artifacts. *Electroencephalography and clinical neurophysiology, 90*(3), 229-241. https://doi.org/10.1016/0013-4694(94)90094-9

Blumstein, S. (2009). Reflections on the Cognitive Neuroscience of Language. In Gazzaniga MS (Ed.), *The Cognitive Neurosciences*. Cambridge, US: MIT Press.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological review, 108*(3), 624. https://doi.org/10.1030//0033-295X.108.3.624

Braver, T. S. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends in cognitive sciences, 16*(2), 106-113. https://doi.org/10.1016/j.tics.2011.12.010.

Brown-Schmidt, S., Yoon, S. O., & Ryskin, R. A. (2015). People as contexts in conversation. *Psychology of learning and motivation,* 62, 59-99. Academic Press. https://doi.org/10.1016/bs.plm.2014.09.003

Brouwer, H., Crocker, M. W., Venhuizen, N. J., & Hoeks, J. C. (2017). A Neurocomputational Model of the N400 and the P600 in Language Processing. *Cognitive science, 41*(S6), 1318-1352. https://doi.org/10.1111/cogs.12461

Carter, Y. D., Lim, S.-J., & Perrachione, T. K. (2019). *Talker continuity facilitates speech processing independent of listeners' expectations.* Paper presented at the 19th International Congress of

Phonetic Sciences. retrieved from https://sites.bu.edu/cnrlab/files/2019/05/Carter-Lim-Perrachione-2019-ICPhS-ERTA.pdf

Chandrasekaran, B., Chan, A. H., & Wong, P. C. (2011). Neural processing of what and who information in speech. *Journal of Cognitive Neuroscience, 23*(10), 2690-2700. https://doi.org/10.1162/jocn.2011.21631

Choi, J. Y., Hu, E. R., & Perrachione, T. K. (2018). Varying acoustic-phonemic ambiguity reveals that talker normalization is obligatory in speech processing. *Attention, Perception, & Psychophysics, 80*(3), 784-797. https://doi.org/10.3758/s13414-017-1395-5

Choi, J. Y., & Perrachione, T. K. (2019). Time and information in perceptual adaptation to speech. *Cognition, 192*, 103982. https://doi.org/10.1016/j.cognition.2019.05.019

Clayson, P. E., & Larson, M. J. (2011). Conflict adaptation and sequential trial effects: Support for the conflict monitoring theory. *Neuropsychologia, 49*(7), 1953-1961. https://doi.org/10.1016/j.neuropsychologia.2011.03.023

Coulson, S. (1998). ERPs and domain specificity: Beating a straw horse. *Language and cognitive processes, 13*(6), 653-672. https://doi.org/10.1080/016909698386410

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and cognitive processes, 13*(1), 21-58. https://doi.org/10.1080/016909698386582

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods, 134*(1), 9-21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proceedings of the National Academy of Sciences, 93*(24), 13494-13499. https://doi.org/10.1073/pnas.93.24.13494

Diamond, A. (2013). Executive Functions. *Annual Review of Psychology, 64*, 135-168. https://doi.org/10.1146/annurev-psych-113011-143750

Dietl, T., Trautner, P., Staedtgen, M., Vannuchi, M., Mecklinger, A., Grunwald, T., . . . Kurthen, M. (2005). Processing of famous faces and medial temporal lobe event-related potentials: a depth electrode study. *Neuroimage, 25*(2), 401-407. https://doi.org/10.1016/j.neuroimage.2004.11.033

Donchin, E. (1981). Surprise!… surprise?. Psychophysiology, 18(5), 493-513. https://doi.org/10.1111/j.1469-8986.1981.tb01815.x

Dörr, P., Herzmann, G., & Sommer, W. (2011). Multiple contributions to priming effects for familiar faces: Analyses with backward masking and event-related potentials. *British Journal of Psychology, 102*(4), 765-782. https://doi.org/10.1111/j.2044-8295.2011.02028.x

Ebner, N. C., Riediger, M., & Lindenberger, U. (2010). FACES—A database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior research methods, 42*(1), 351-362. https://doi.org/10.3758/BRM.42.1.351

Eimer, M. (1998). Does the face-specific N170 component reflect the activity of a specialized eye processor? *NeuroReport, 9*(13), 2945-2948. https://doi.org/10.1097/00001756-199809140-00005

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive Brain Research, 1*(3), 183-192. https://doi.org/10.1016/0926-6410(93)90026-2

Goslin, J., Duffy, H., & Floccia, C. (2012). An ERP investigation of regional and foreign accent processing. *Brain Lang, 122*(2), 92-102. doi:10.1016/j.bandl.2012.04.017

Gratton, G., Cooper, P., Fabiani, M., Carter, C. S., & Karayanidis, F. (2018). Dynamics of cognitive control: Theoretical bases, paradigms, and a view for the future. *Psychophysiology, 55*(3), e13016. https://doi.org/10.1111/psyp.13016

Grey, S., & van Hell, J. G. (2017). Foreign-accented speaker identity affects neural correlates of language comprehension. *Journal of Neurolinguistics, 42*, 93-108. doi:10.1016/j.jneuroling.2016.12.001

Grey, S., Cosgrove, A. L., & van Hell, J. G. (2020). Faces with foreign accents: An event-related potential study of accented sentence comprehension. *Neuropsychologia*, *147*, 107575. https://doi.org/10.1016/j.neuropsychologia.2020.107575

Grill-Spector, K., & Malach, R. (2001). fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta psychologica, 107*(1-3), 293-321. https://doi.org/10.1016/S0001-6918(01)00019-1

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in cognitive sciences, 10*(1), 14-23. https://doi.org/10.1016/j.tics.2005.11.006

Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. *Journal of Cognitive Neuroscience, 11*(2), 194-205. https://doi.org/10.1162/089892999563328

Hanulíková, A., van Alphen, P. M., van Goch, M. M., & Weber, A. (2012). When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience, 24*(4), 878-887. https://doi.org/10.1162/jocn_a_00103

Hernández-Gutiérrez, D., Muñoz, F., Sánchez-García, J., Sommer, W., Abdel Rahman, R., Casado, P., ... & Martín-Loeches, M. (2021). Situating language in a minimal social context: how seeing a picture of the speaker’s face affects language comprehension. *Social Cognitive and Affective Neuroscience*. https://doi.org/10.1093/scan/nsab009

Herzmann, G., Schweinberger, S. R., Sommer, W., & Jentzsch, I. (2004). What's special about personally familiar faces? A multimodal approach. *Psychophysiology, 41*(5), 688-701. https://doi.org/10.1111/j.1469-8986.2004.00196.x

Jacques, C., & Rossion, B. (2006). The speed of individual face categorization. *Psychological science, 17*(6), 485-492. https://doi.org/10.1111/j.1467-9280.2006.01733.x

Jacques, C., d'Arripe, O., & Rossion, B. (2007). The time course of the inversion effect during individual face discrimination. *Journal of Vision*, 7(8), 3-3. https://doi.org/10.1167/7.8.3

Johnson, R. (1986). A triarchic model of P300 amplitude. *Psychophysiology*. https://doi.org/10.1111/j.1469-8986.1986.tb00649.x

Jost, K., Mayr, U., & Rösler, F. (2008). Is task switching nothing but cue priming? Evidence from ERPs. *Cognitive, Affective, & Behavioral Neuroscience*, *8*(1), 74-84. https://doi.org/10.3758/CABN.8.1.74

Kaan, E., & Swaab, T. Y. (2003). Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of Cognitive Neuroscience, 15*(1), 98-110. https://doi.org/10.1162/089892903321107855

Kaganovich, N., Francis, A. L., & Melara, R. D. (2006). Electrophysiological evidence for early interaction between talker and linguistic information during speech perception. *Brain research, 1114*(1), 161-172. https://doi.org/10.1016/j.brainres.2006.07.049

Kapadia, A. M., & Perrachione, T. K. (2020). Selecting among competing models of talker adaptation: Attention, cognition, and memory in speech processing efficiency. *Cognition, 204*, 104393. https://doi.org/10.1016/j.cognition.2020.104393

Key-DeLyria, S. E., & Altmann, L. J. (2016). Executive function and ambiguous sentence comprehension. *American journal of speech-language pathology, 25*(2), 252-267. https://doi.org/10.1044/2015_AJSLP-14-0111

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching — A review. *Psychological bulletin*, *136*(5), 849. https://doi.org/10.1037/a0019842

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological review, 122*(2), 148. https://doi.org/10.1037/a0038695

Kolk, H., & Chwilla, D. (2007). Late positivities in unusual situations. *Brain and language, 100*(3), 257-261. doi:10.1016/j.bandl.2006.07.006

Kopp, B., Steinke, A., & Visalli, A. (2020). Cognitive flexibility and N2/P3 event-related brain potentials. *Scientific reports, 10*(1), 1-17.https://doi.org/10.1038/s41598-020-66781-5

Kroczek, L. O. H., & Gunter, T. C. (2017). Communicative predictions can overrule linguistic priors. *Sci Rep, 7*(1), 17581. doi:10.1038/s41598-017-17907-9

Kroczek, L. O., & Gunter, T. C. (2020). Distinct neural networks relate to common and speaker-specific language priors. *Cerebral Cortex Communications*, *1*(1), tgaa021. https://doi.org/10.1093/texcom/tgaa021

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension?. *Language, cognition and neuroscience*, *31*(1), 32-59. https://doi.org/10.1080/23273798.2015.1102299

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*(4427), 203-205. https://doi.org/10.1126/science.7350657

Leckey, M., & Federmeier, K. D. (2020). The P3b and P600(s): Positive contributions to language comprehension. *Psychophysiology, 57*(7), e13351. https://doi.org/10.1111/psyp.13351

Li, L., Miller, E. K., & Desimone, R. (1993). The representation of stimulus familiarity in anterior inferior temporal cortex. *Journal of neurophysiology, 69*(6), 1918-1929. https://doi.org/10.1152/jn.1993.69.6.1918

Lim, S.-J., Shinn-Cunningham, B. G., & Perrachione, T. K. (2019). Effects of talker continuity and speech rate on auditory working memory. *Attention, Perception, & Psychophysics, 81*(4), 1167-1177. https://doi.org/10.3758/s13414-019-01684-w

Lim, S.-J., Tin, J. A., Qu, A., & Perrachione, T. K. (2019). *Attentional reorientation explains processing costs associated with talker variability.* Paper presented at the 19th International Congress of Phonetic Sciences. retrieved from http://sites.bu.edu/cnrlab/files/2019/05/Lim-Tin-Qu-Perrachione-2019-ICPhS-TCTA.pdf

Lindenberger, U., Ebner, N., & Riediger, M. (2005-2007). FACES collection, from Max Planck Institute for Human Development, Center for Lifespan Psychology, Berlin, Germany website: http://faces.mpib-berlin.mpg.de/album/escidoc:57488

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in psychology*, *4*, 226. https://doi.org/10.3389/fpsyg.2013.00226

Magnuson, J. S., & Nusbaum, H. C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology: Human Perception and Performance, 33*(2), 391. https://doi.org/10.1037/0096-1523.33.2.391

Miller, E. K., & Desimone, R. (1994). Parallel neuronal mechanisms for short-term memory. *Science, 263*(5146). https://doi.org/520-522. 10.1126/science.8290960

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49-100. https://doi.org/10.1006/cogp.1999.0734

Münte, T. F., Heinze, H.-J., Matzke, M., Wieringa, B. M., & Johannes, S. (1998). Brain potentials and syntactic violations revisited: No evidence for specificity of the syntactic positive shift. *Neuropsychologia, 36*(3), 217-226. https://doi.org/10.1016/S0028-3932(97)00119-X

Nusbaum, H. C., & Morin, T. M. (1992). Paying attention to differences among talkers. In Y. Tohkura, E. Vatikiotis-Bateson, Y. Sagisaka (Ed.), *Speech perception, production and linguistic structure* (pp. 113-134). Tokyo: Ohmsha, Ltd.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia, 9*(1), 97-113. https://doi.org/10.1016/0028-3932(71)90067-4

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational intelligence and neuroscience, 2011*, 1. https://doi.org/10.1155/2011/156869

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31*(6), 785-806. https://doi.org/10.1016/0749-596X(92)90039-Z

Regel, S., Coulson, S., & Gunter, T. C. (2010). The communicative style of a speaker can affect language comprehension? ERP evidence from the comprehension of irony. *Brain research, 1311*, 121-135. https://doi.org//10.1016/j.brainres.2009.10.077

Romero-Rivas, C., Martin, C. D., & Costa, A. (2015). Processing changes when listening to foreign-accented speech. *Front Hum Neurosci*, 9, 167. doi:10.3389/fnhum.2015.00167

Rossion, B. (2014). Understanding face perception by means of human electrophysiology. *Trends in cognitive sciences*, 18(6), 310-318. https://doi.org/10.1016/j.tics.2014.02.013

Rossion, B., & Jacques, C. (2011). The N170: understanding the time-course of face perception in the human brain. In S. J. Luck, & E. S. Kappenman (Eds.), *The Oxford handbook of ERP components* (115-142). Oxford: Oxford University Press.

Sassenhagen, J., & Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology*, e13335. https://doi.org/10.1111/psyp.13335

Sassenhagen, J., & Fiebach, C. J. (2019). Finding the P3 in the P600: Decoding shared neural mechanisms of responses to syntactic violations and oddball targets. *NeuroImage, 200*, 425-436. https://doi.org/10.1016/j.neuroimage.2019.06.048

Sassenhagen, J., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2014). The P600-as-P3 hypothesis revisited: Single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned. *Brain and language, 137*, 29-39. https://doi.org/10.1016/j.bandl.2014.07.010

Schweinberger, S. R., Pfütze, E.-M., & Sommer, W. (1995). Repetition priming and associative priming of face recognition: Evidence from event-related potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(3), 722. https://doi.org/10.1037/0278-7393.21.3.722

Schweinberger, S. R., Pickering, E. C., Jentzsch, I., Burton, A. M., & Kaufmann, J. M. (2002). Event-related brain potential evidence for a response of inferior temporal cortex to familiar face repetitions. *Cognitive Brain Research*, 14(3), 398-409. https://doi.org/10.1016/s0926-6410(02)00142-8

Schweinberger, S. R., & Neumann, M. F. (2016). Repetition effects in human ERPs to faces. *Cortex, 80*, 141-153. https://doi.org/10.1016/j.cortex.2015.11.001

Shinn-Cunningham, B.G. 2008. Object-based auditory and visual attention. Trends Cogn. Sci. 12, 182-186. https://doi.org/10.1016/j.tics.2008.02.003

van Berkum, J. J., van den Brink, D., Tesink, C. M., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience, 20*(4), 580-591. https://doi.org/10.1162/jocn.2008.20054

van de Meerendonk, N., Indefrey, P., Chwilla, D. J., & Kolk, H. H. (2011). Monitoring in language perception: Electrophysiological and hemodynamic responses to spelling violations. *Neuroimage, 54*(3), 2350-2363. https://doi.org/10.1016/j.neuroimage.2010.10.022

van de Meerendonk, N., Kolk, H. H., Chwilla, D. J., & Vissers, C. T. W. (2009). Monitoring in language perception. *Language and Linguistics Compass, 3*(5), 1211-1224. https://doi.org/10.1111/j.1749-818X.2009.00163.x

van Herten, M., Kolk, H. H., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Brain Res Cogn Brain Res, 22*(2), 241-255. doi:10.1016/j.cogbrainres.2004.09.002

von Kriegstein, K., Dogan, O., Grüter, M., Giraud, A.-L., Kell, C. A., Grüter, T., Kleinschmidt, A., Kiebel, S. J. (2008). Simulation of talking faces in the human brain improves auditory speech recognition. *Proceedings of the National Academy of Sciences, 105*(18), 6747-6752. https://doi.org/10.1073/pnas.0710826105

Vissers, C. T. W., Kolk, H. H., van de Meerendonk, N., & Chwilla, D. J. (2008). Monitoring in language perception: evidence from ERPs in a picture–sentence matching task. *Neuropsychologia, 46*(4), 967-982. https://doi.org/10.1016/j.neuropsychologia.2007.11.027

Weissgerber, T. L., Milic, N. M., Winham, S. J., & Garovic, V. D. (2015). Beyond bar and line graphs: time for a new data presentation paradigm. *PLoS Biol, 13*(4), e1002128. https://doi.org/10.1371/journal.pbio.1002128

Wiese, H., Tüttenberg, S. C., Ingram, B. T., Chan, C. Y., Gurbuz, Z., Burton, A. M., & Young, A. W. (2019). A robust neural index of high face familiarity. Psychological *science*, 30(2), 261-272. https://doi.org/10.1177/0956797618813572

Wiggs, C. L., & Martin, A. (1998). Properties and mechanisms of perceptual priming. *Current opinion in neurobiology, 8*(2), 227-233. https://doi.org/10.1016/s0959-4388(98)80144-x

Wong, P. C., Nusbaum, H. C., & Small, S. L. (2004). Neural bases of talker normalization. *Journal of Cognitive Neuroscience, 16*(7), 1173-1184. https://doi.org/10.1162/0898929041920522

Xu, J., Abdel Rahman, R., & Sommer, W. (2019). Perceived language competence modulates criteria for speech error processing: evidence from event-related potentials. *Language, Cognition and Neuroscience*, 1-14. https://doi.org/10.1080/23273798.2018.1562558

Xu, J., Abdel Rahman, R., & Sommer, W. (2021). Sequential adaptation effects reveal proactive control in processing spoken sentences: Evidence from event-related potentials. *Brain and Language*, 214, 104904. https://doi.org/10.1016/j.bandl.2020.104904

Author Notes

The authors report no conflict of interest. We thank the reviewers for their valuable comments.

This work was supported by the China Scholarship Council (CSC) under Grant [201606230215] to J.X.. The funder had no role in data analysis, decision to publish, or preparation of the manuscript.

Footnotes

[1] In the present analyses, the P600 time window was determined based on a previous study (Xu et al., 2019), because the present study used the same experimental materials and similar presentation paradigms as the previous study and we had very specific expectations regarding the P600 component. Out of the concern that potential findings might be found if using more sensitive analysis approaches, for example, cluster-based permutation tests, we have also applied cluster-based permutation tests to explore speaker identity effects in Block 1 and 2 separately. The cluster-based permutation tests found that the 10%-speaker condition had P600 effects in both blocks, the 50%-speaker only in Block 2, and the 90%-speaker no P600 effects in either block, which were similar to the current ANOVA results. But we finally decided to stick just with previously-determined P600 ROI and time window and using ANOVA because of the clear hypotheses and for the sake of comparability with our previous studies (Xu et al., 2019, 2021). Furthermore, just using permutation tests to establish significance of effect latency or location might be problematic (see for example Sassenhagen & Draschkow, 2019 for a discussion on this issue).

[2] In order to provide more information about the impact of the outlier on results, we conducted the same main analyses using all data sets including the outlier, namely, the speaker sequence analysis, the speaker identity analysis, and face ERPs analysis. All statistical results can be found in the Supplement. Briefly, the only difference between results including and excluding the outlier was that in the speaker sequence analysis the interaction between block and speaker sequence was a trend when including the outlier and significant without the outlier; both pairwise comparisons showed that the speaker sequence effect was significant in Block 1 and nonsignificant in Block 2. This indicates that including the outlier had little impact on the P600 results, and that it was not the exclusion of the outlier that resulted in finding the speaker sequence and face ERP effects.

[3] It is unlikely that this large effect was merely an artefact caused by the unequal trial numbers between the correct and incorrect sentence conditions for the 10%-speaker; in this case, a similarly strong negative effect should have been generated for the 90%-speaker, who had exactly the reversed trial numbers in the two sentence conditions, which was not observed. Nevertheless, due to the trial

number issue and the exploratory nature of this analysis, this trend in differential P600 effects between the 10%-speaker and the other two must be interpreted with caution.

Tables

Table 1: Sentence Examples

Notes: Adapted from Appendix C in Xu et al. (2019). In each example, a. is the well-formed version, and b. is the version with a grammatical agreement violation. English translations of the well-formed German sentences are presented in round brackets. The grammatical genders are given as subscripts in square brackets (m = masculine, f = feminine, n = neuter). If the grammatical gender is subscripted under nouns, it refers to the grammatical gender of this noun; if subscripted under the determiners before the nouns, it refers to the correct grammatical gender that the determiner should lead. The subscripts $_{[singular]}$ and $_{[plural]}$ indicate the grammatical number of the subscripted word. The case of the noun or the case that the verb should govern is given in square brackets as subscripts (N = nominative, G = genitive, D = dative, A = accusative).

(1) *Determiner-Noun Gender Agreement*

a. Ich hätte gerne einen$_{[m]}$ Mocca$_{[m]}$ mit Sahne.

b. Ich hätte gerne eine$_{[f]}$ *Mocca*$_{[m]}$ mit Sahne.

(I would like a$_{[m]}$ mocha$_{[m]}$ with cream.)

(2) *Subject/Pronoun-Verb Number Agreement*

a. Er$_{[singular]}$ verließ$_{[singular]}$ entsetzt das Büro.

b. Er$_{[singular]}$ *verließen*$_{[plural]}$ entsetzt das Büro.

(He$_{[singular]}$ *left*$_{[singular]}$ the office in shock.)

(3) *Determiner-Noun Number Agreement*

a. Stell die Nudeln zwei$_{[plural]}$ Minuten$_{[plural]}$ in die Mikrowelle.

b. Stell die Nudeln zwei$_{[plural]}$ *Minute*$_{[singular]}$ in die Mikrowelle.

(Put the noodles in the microwave for two$_{[plural]}$ minutes$_{[plural]}$.)

(4) *Verb-Object Case Agreement*

a. Willst du mich$_{[A]}$ erstechen$_{[A]}$, pass doch auf!

b. Willst du *mir*$_{[D]}$ erstechen$_{[A]}$, pass doch auf!

(Do you want to stab$_{[A]}$ *me*$_{[A]}$, be careful!)

Table 2: Speaker Sequence Analysis Results

	df	*F*	*p*	η_p^2
speaker sequence	1, 37	3.081	.088	.077
correctness	1, 37	11.494	.002	.237
block	1, 37	9.976	.003	.212
correctness * block	1, 37	.088	.354	.023
speaker sequence * correctness	1, 37	.648	.426	.017
speaker sequence * block	1, 37	5.341	.027	.126
speaker sequence * correctness * block	1, 37	.093	.762	.003

Pairwise comparison on the significant interaction between speaker sequence and block:

Block 1	*df*	*F*	*p*	η_p^2
speaker sequence	1, 37	7.783	.008	.174
Block 2	*df*	*F*	*p*	η_p^2
speaker sequence	1, 37	.002	.964	.000

Table 3: Speaker Identity Analysis Results

Block 2	***df***	***F***	***p***	η_p^2
speaker identity	2, 74	.608	.547	.016
correctness	1, 37	10.609	.002	.223
speaker identity * correctness	2, 74	.164	.849	.004

Table 4: Speaker-specific Accuracy

***Mean* (*SD*)**	**Block 1**	**Block 2**	**Overall**
90%-speaker	84.49% (9.71%)	88.85% (8.51%)	86.67% (9.11%)
50%-speaker	88.73% (7.13%)	91.14% (7.14%)	89.94% (7.14%)
10%-speaker	93.01% (6.25%)	88.57% (9.41%)	90.79% (7.83%)

Table 5: Exploratory Analysis on Speaker-identity Effect in Block 1

Block 1	***df***	***F***	***p***	η_p^2
speaker identity	2, 74	1.483	.234	.039
correctness	1, 37	1.517	.226	.039
speaker identity*correctness	2, 74	2.539	.086	.064

Pairwise comparison on the interaction between speaker identity and correctness:

correctness in Block 1	***df***	***F***	***p***	η_p^2
90%-speaker	1, 37	.134	.717	.004
50%-speaker	1, 37	.190	.665	.005
10%-speaker	1, 37	4.819	.034	.115

Table 6: Face ERPs Analysis Results

N170	***df***	***F***	***p***	η_p^2
hemisphere	1, 37	.514	.478	.014
block	1, 37	.499	.484	.013
speaker sequence	1, 37	19.756	< .001	.348
hemisphere * block	1, 37	.370	.547	.010
block * speaker sequence	1, 37	.006	.937	.000
hemisphere * speaker sequence	1, 37	1.065	.309	.028
hemisphere * block * speaker sequence	1, 37	.048	.827	.001
N250r	***df***	***F***	***p***	η_p^2
hemisphere	1, 37	12.552	.001	.253
block	1, 37	.692	.411	.018
speaker sequence	1, 37	60.348	< .001	.620
hemisphere * block	1, 37	.351	.557	.009
block * speaker sequence	1, 37	11.047	.002	.230
hemisphere * speaker sequence	1, 37	1.743	.195	.045
hemisphere * block * speaker sequence	1, 37	.494	.486	.013

Pairwise comparison on the significant interaction between speaker sequence and block:

a.

Speaker sequence	***df***	***F***	***p***	η_p^2
Block 1	1, 37	64.225	< .001	.634
Block 2	1, 37	41.264	< .001	.527

b.

Block	***df***	***F***	***p***	η_p^2
Speaker repetition	1, 37	.309	.581	.008
Speaker switch	1, 37	3.364	.075	.083

Figures

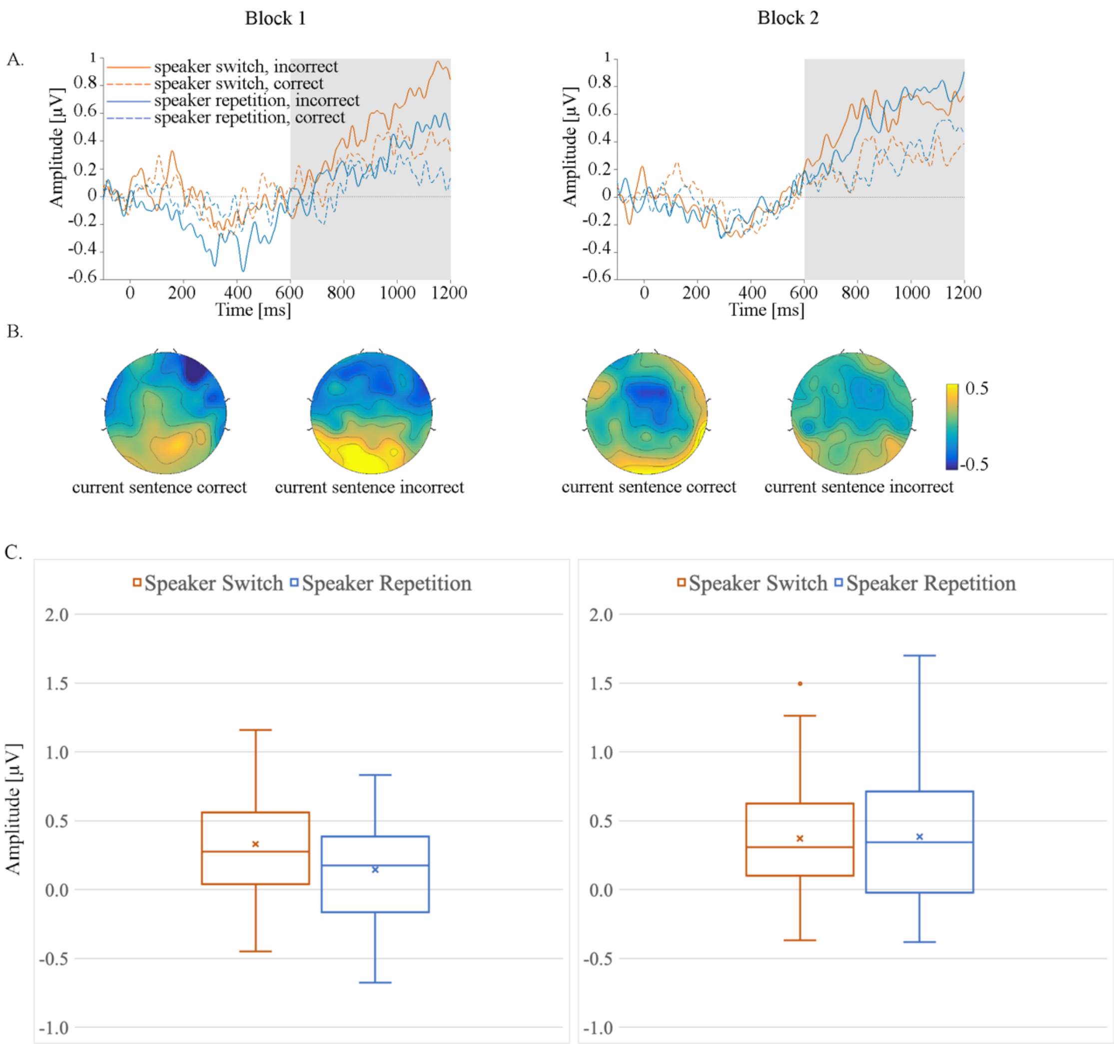


Figure 1. Speaker sequence analyses. **A.** ERPs represent grand means ($N = 38$) in P600-ROI, averaged separately for each speaker sequence and sentence correctness condition in Block 1 and 2, time-locked to onsets of critical words in target sentences. Positive is plotted upward. Time window for the P600 effect is shaded. **B.** Difference topographies of P600 segments (grand means of 38 participants) for speaker switch minus speaker repetition (600-1200 ms), averaged separately for correct and incorrect current sentence conditions in Block 1 and 2. **C.** The boxplots display all participants' ($N = 38$) mean P600 amplitudes, separately averaged for each participant in the two speaker sequence conditions in Block 1 and 2, respectively.

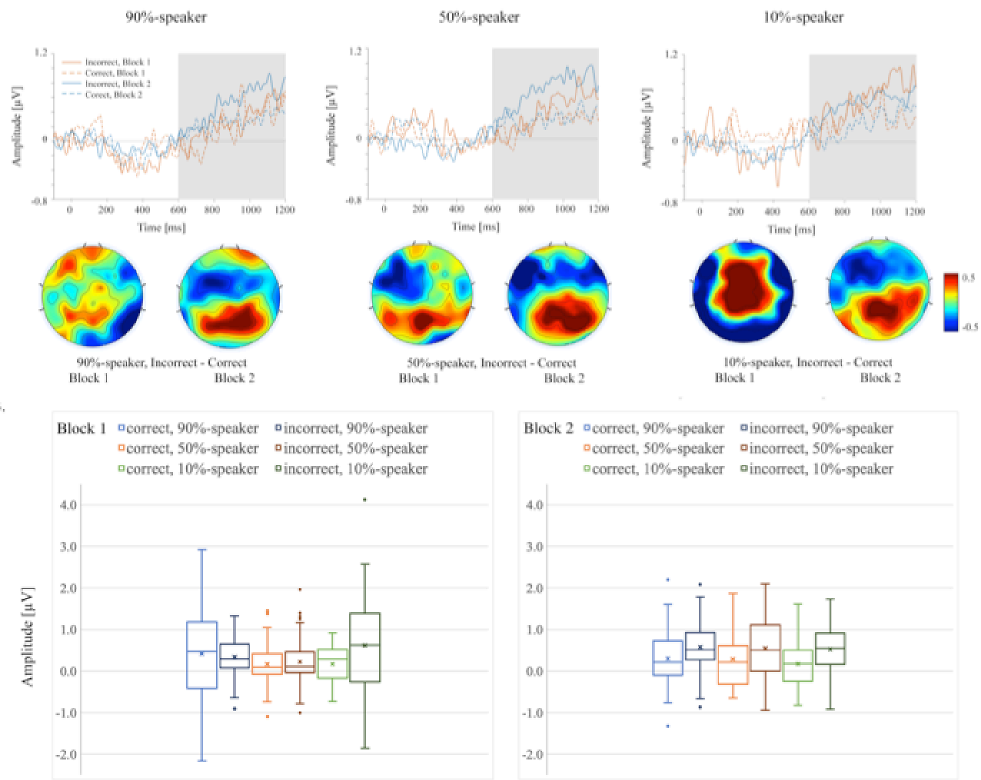


Figure 2. Speaker identity analyses. **A.** ERPs represent grand means ($N = 38$) in P600-ROI, averaged separately for each speaker identity and sentence correctness condition in Block 1 and 2, time-locked to onsets of critical words in target sentences. Positive is plotted upward. Time window for the P600 effect is shaded. Difference topographies of P600 segments (grand means of 38 participants) for incorrect minus correct sentence conditions (600-1200 ms), averaged separately for each speaker identity in Block 1 and 2. **B.** The boxplots display all participants' ($N = 38$) mean P600-ROI amplitudes, time-locked to onsets of critical words in target sentences, averaged separately for each participant in each speaker identity and sentence correctness condition in Block 1 and 2, respectively.

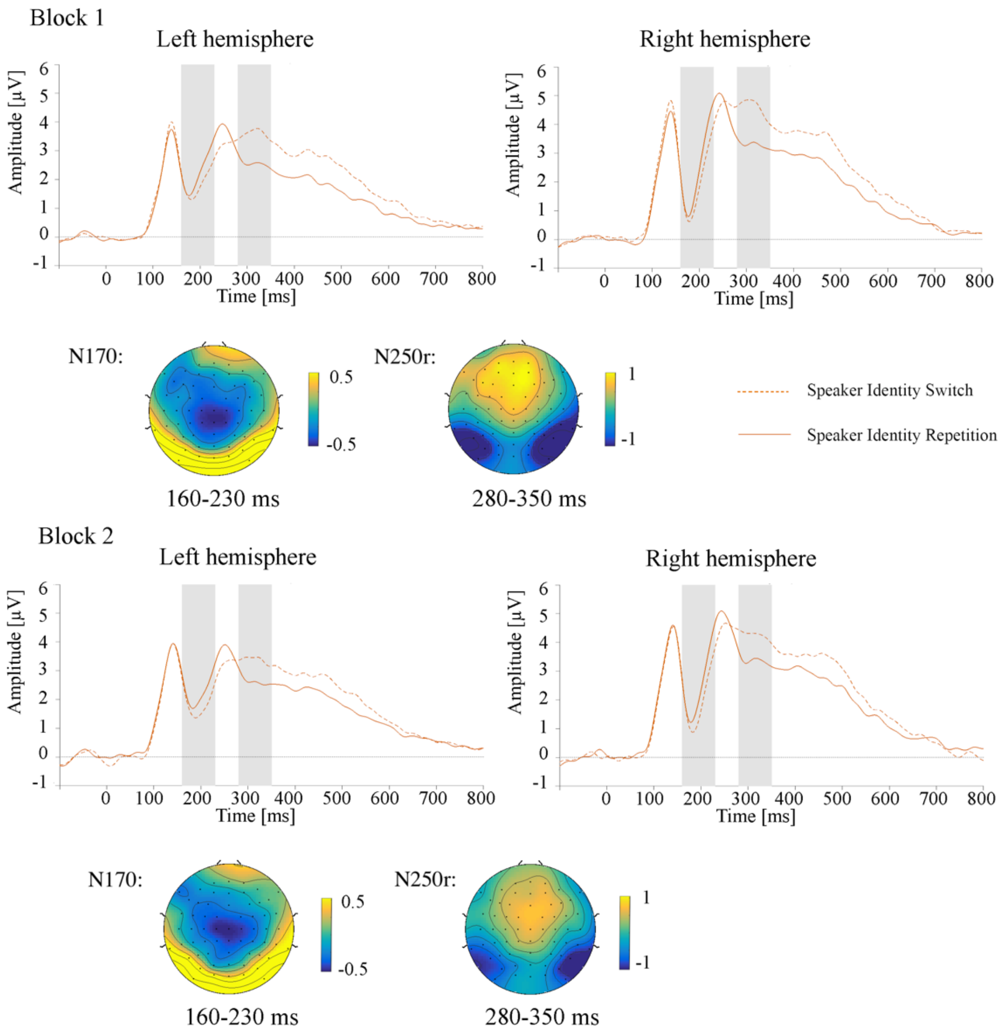


Figure 3: Face-elicited ERPs. ERP waveforms in N170 and N250 ROI, time-locked to onsets of cue faces preceding target sentences in Block 1 and 2. Shaded areas mark the time windows of N170 (160-230 ms) and N250 (280-350 ms). Difference topographies represent difference maps of speaker repetition minus speaker switch separately averaged for N170 and N250 time windows in Block 1 and 2.

Declaration

This work was supported by a doctoral scholarship from the China Scholarship Council (CSC) under Grant [201606230215] to J.X.. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

No potential conflict of interest was reported by the authors.

We thank many colleagues and student assistants for helping with data collection.

Correspondence concerning this article should be sent to: Jue Xu, Department of Psychology, Humboldt-Universität zu Berlin, Rudower Chaussee 18, 12489 Berlin, Germany, xujue@hu-berlin.de.

Complete Reference List

Alpay, G., Goerke, M., & Stürmer, B. (2009). Precueing imminent conflict does not override sequence-dependent interference adaptation. *Psychological Research PRPF*, *73*(6), 803-816. doi: 10.1007/s00426-008-0196-9

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentais, prosody, and syllable structure. *Language learning*, 42(4), 529-555. doi: 10.1111/j.1467-1770.1992.tb01043.x

Barber, H., & Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: An ERP comparison. Journal of Cognitive Neuroscience, 17(1), 137-153

Barrett, S. E., Rugg, M. D., & Perrett, D. I. (1988). Event-related potentials and the matching of familiar and unfamiliar faces. Neuropsychologia, 26(1), 105-117. doi: 10.1016/0028-3932(88)90034-6

Barrett, S., & Rugg, M. D. (1989). Event-related potentials and the semantic matching of faces. Neuropsychologia, 27(7), 913-922. doi: 10.1016/0028-3932(89)90067-5

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. Journal of Cognitive Neuroscience, 8(6), 551-565. doi: 10.1162/jocn.1996.8.6.551

Berg, P., & Scherg, M. (1994). A multiple source approach to the correction of eye artifacts. *Electroencephalography and clinical neurophysiology*, 90(3), 229-241. doi: 10.1016/0013-4694(94)90094-9

Berry, G. M. (2016). Processing linguistic variation through dual mechanisms of cognitive control. Linguistics Vanguard, 2(s1). doi: 10.1515/lingvan-2016-0033

Blumstein, S. (2009). Reflections on the Cognitive Neuroscience of Language. In Gazzaniga MS (Ed.), *The Cognitive Neurosciences*. Cambridge, US: MIT Press

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological review*, 108(3), 624. doi: 10.1030//0033-295X.108.3.624

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: an update. *Trends in cognitive sciences*, *8*(12), 539-546. doi: 10.1016/j.tics.2004.10.003

Braver, T. S. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends in cognitive sciences*, 16(2), 106-113. doi: 10.1016/j.tics.2011.12.010

Braver, T. S., Gray, J. R., & Burgess, G. C. (2007). Explaining the many varieties of working memory variation: Dual mechanisms of cognitive control. In A. R. A. Conway, C. Jarrold, M. J. Kane (Eds.) & A. Miyake & J. N. Towse (Ed.), *Variation in working memory* (pp. 76-106). New York, US: Oxford University Press

Brouwer, H., Crocker, M. W., Venhuizen, N. J., & Hoeks, J. C. (2017). A Neurocomputational Model of the N400 and the P600 in Language Processing. *Cognitive science*, 41(S6), 1318-1352. doi: 10.1111/cogs.12461

Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about semantic illusions: rethinking the functional role of the P600 in language comprehension. *Brain research*, *1446*, 127-143. doi: 10.1016/j.brainres.2012.01.055

Burgess, G. C., & Braver, T. S. (2010). Neural mechanisms of interference control in working memory: effects of interference expectancy and fluid intelligence. *PLoS One*, 5(9), e12861. doi: 10.1371/journal.pone.0012861

Campbell, K. L., & Tyler, L. K. (2018). Language-related domain-specific and domain-general systems in the human brain. Current opinion in behavioral sciences, 21, 132-137. doi: 10.1016/j.cobeha.2018.04.008

Carter, Y. D., Lim, S.-J., & Perrachione, T. K. (2019). Talker continuity facilitates speech processing independent of listeners' expectations. Paper presented at the 19th International Congress of Phonetic Sciences. retrieved from https://sites.bu.edu/cnrlab/files/2019/05/Carter-Lim-Perrachione-2019-ICPhS-ERTA.pdf

Chandrasekaran, B., Chan, A. H., & Wong, P. C. (2011). Neural processing of what and who information in speech. *Journal of Cognitive Neuroscience*, 23(10), 2690-2700. doi: 10.1162/jocn.2011.21631

Chen, L., Shu, H., Liu, Y., Zhao, J., & Li, P. (2007). ERP signatures of subject–verb agreement in L2 learning. Bilingualism: Language and Cognition, 10(2), 161-174

Choi, J. Y., & Perrachione, T. K. (2019). Time and information in perceptual adaptation to speech. Cognition, 192, 103982. doi: 10.1016/j.cognition.2019.05.019

Choi, J. Y., Hu, E. R., & Perrachione, T. K. (2018). Varying acoustic-phonemic ambiguity reveals that talker normalization is obligatory in speech processing. Attention, Perception, & Psychophysics, 80(3), 784-797. doi: 10.3758/s13414-017-1395-5

Clayson, P. E., & Larson, M. J. (2011). Conflict adaptation and sequential trial effects: Support for the conflict monitoring theory. *Neuropsychologia*, *49*(7), 1953-1961. doi: 10.1016/j.neuropsychologia.2011.03.023

Columbus, G., Sheikh, N. A., Côté-Lecaldare, M., Häuser, K., Baum, S. R., & Titone, D. (2015). Individual differences in executive control relate to metaphor processing: an eye movement study of sentence reading. Frontiers in human neuroscience, 8, 1057. doi: 10.3389/fnhum.2014.01057

Connolly, J. F., & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. Journal of Cognitive Neuroscience, 6(3), 256-266

Coulson, S. (1998). ERPs and domain specificity: Beating a straw horse. *Language and cognitive processes*, 13(6), 653-672. doi: 10.1080/016909698386410

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and cognitive processes*, 13(1), 21-58. doi: 10.1080/016909698386582

Cristia, A., Seidl, A., Vaughn, C., Schmale, R., Bradlow, A., & Floccia, C. (2012). Linguistic processing of accented speech across the lifespan. Frontiers in psychology, 3, 479

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, 134(1), 9-21. doi: 10.1016/j.jneumeth.2003.10.009

Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. Proceedings of the National Academy of Sciences, 93(24), 13494-13499. doi: 10.1073/pnas.93.24.13494

Diamond, A. (2013). Executive Functions. Annual Review of Psychology, 64, 135-168. doi: 10.1146/annurev-psych-113011-143750

Dietl, T., Trautner, P., Staedtgen, M., Vannuchi, M., Mecklinger, A., Grunwald, T., . . . Kurthen, M. (2005). Processing of famous faces and medial temporal lobe event-related potentials: a depth electrode study. Neuroimage, 25(2), 401-407. doi: 10.1016/j.neuroimage.2004.11.033

Dörr, P., Herzmann, G., & Sommer, W. (2011). Multiple contributions to priming effects for familiar faces: Analyses with backward masking and event-related potentials. British Journal of Psychology, 102(4), 765-782. doi: 10.1111/j.2044-8295.2011.02028.x

Donchin, E., & Coles, M. G. (1988). Is the P300 component a manifestation of context updating. *Behavioral and brain sciences*, *11*(3), 357-427

Donchin, E. (1981). Surprise!… surprise?. *Psychophysiology*, *18*(5), 493-513

Dreisbach, G., & Fischer, R. (2012). Conflicts as aversive signals. *Brain and cognition*, 78(2), 94-98. doi: 10.1016/j.bandc.2011.12.003

Ebner, N. C., Riediger, M., & Lindenberger, U. (2010). FACES—A database of facial expressions in young, middle-aged, and older women and men: Development and validation. Behavior research methods, 42(1), 351-362. doi: 10.3758/BRM.42.1.351

Eimer, M. (1998). Does the face-specific N170 component reflect the activity of a specialized eye processor? NeuroReport, 9(13), 2945-2948. doi: 10.1097/00001756-199809140-00005

Fedorenko, E. (2014). The role of domain-general cognitive control in language comprehension. Frontiers in psychology, 5, 335. doi: 10.3389/fpsyg.2014.00335

Franceschina, F. (2005). Fossilized second language grammars: The acquisition of grammatical gender. *Language Acquisition and Language Disorders* (Vol. 38, xxiv, 288 pp.). Lancaster, UK: John Benjamins Publishing

Fridlund, A. J., & Cacioppo, J. T. (1986). Guidelines for human electromyographic research. Psychophysiology, 23(5), 567-589

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive Brain Research*, 1(3), 183-192. doi: 10.1016/0926-6410(93)90026-2

Gao, W., Cao, B., Shan, S., Chen, X., Zhou, D., Zhang, X., & Zhao, D. (2008). The CAS-PEAL large-scale Chinese face database and baseline evaluations. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 38(1), 149-161

Gonthier, C., Braver, T. S., & Bugg, J. M. (2016). Dissociating proactive and reactive control in the Stroop task. Memory & Cognition, 44(5), 778-788. doi: 10.3758%2Fs13421-016-0591-1

Goslin, J., Duffy, H., & Floccia, C. (2012). An ERP investigation of regional and foreign accent processing. *Brain Lang*, 122(2), 92-102. doi:10.1016/j.bandl.2012.04.017

Gratton, G., Coles, M. G., & Donchin, E. (1992). Optimizing the use of information: strategic control of activation of responses. *Journal of Experimental Psychology*: General, 121(4), 480. doi: 10.1037/0096-3445.121.4.480

Gratton, G., Cooper, P., Fabiani, M., Carter, C. S., & Karayanidis, F. (2018). Dynamics of cognitive control: Theoretical bases, paradigms, and a view for the future. *Psychophysiology*, 55(3), e13016. doi: 10.1111/psyp.13016

Green, K. P., Tomiak, G. R., & Kuhl, P. K. (1997). The encoding of rate and talker information during phonetic perception. *Perception & Psychophysics*, 59(5), 675-692. doi: 10.3758/BF03206015

Grey, S., Cosgrove, A. L., & van Hell, J. G. (2020). Faces with foreign accents: An event-related potential study of accented sentence comprehension. *Neuropsychologia, 147*, 107575.doi: 10.1016/j.neuropsychologia.2020.107575

Grey, S., & van Hell, J. G. (2017). Foreign-accented speaker identity affects neural correlates of language comprehension. *Journal of Neurolinguistics*, 42, 93-108. doi: 10.1016/j.jneuroling.2016.12.001

Grill-Spector, K., & Malach, R. (2001). fMR-adaptation: a tool for studying the functional properties of human cortical neurons. Acta psychologica, 107(1-3), 293-321. doi: 10.1016/S0001-6918(01)00019-1

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. Trends in cognitive sciences, 10(1), 14-23. doi: 10.1016/j.tics.2005.11.006

Gunter, T. C., & Friederici, A. D. (1999). Concerning the automaticity of syntactic processing. Psychophysiology, 36(1), 126-137. doi: 10.1017/S004857729997155X

Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. Journal of Cognitive Neuroscience, 15(6), 883-899

Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech compared to reading: the P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. Neuropsychologia, 38(11), 1531-1549

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. Language and cognitive processes, 8(4), 439-483. doi: 10.1080/01690969308407585

Hahne, A., & Friederici, A. D. (1998). ERP-evidence for autonomous first-pass parsing processes in auditory language comprehension. *Journal of Cognitive Neuroscience, Supplement, 125*

Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. *Journal of Cognitive Neuroscience*, 11(2), 194-205. doi: 10.1162/089892999563328

Hanulíková, A., van Alphen, P. M., van Goch, M. M., & Weber, A. (2012). When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience*, 24(4), 878-887. doi: 10.1162/jocn_a_00103

Heister, J., Würzner, K.-M., Bubenzer, J., Pohl, E., Hanneforth, T., Geyken, A., & Kliegl, R. (2011). dlexDB – eine lexikalische Datenbank für die psychologische und linguistische

Forschung [dlexDB – a lexical data bank for psychological and linguistic researches]. Psychologische Rundschau

Hernández-Gutiérrez, D., Abdel Rahman, R., Martín-Loeches, M., Muñoz, F., Schacht, A., & Sommer, W. (2018). Does dynamic information about the speaker's face contribute to semantic speech processing? ERP evidence. Cortex, 104, 12-25. doi: 10.1016/j.cortex.2018.03.031

Hernández-Gutiérrez, D., Muñoz, F., Sánchez-García, J., Sommer, W., Abdel Rahman, R., Casado, P., Jiménez-Ortega, L., Espuny J., Fondevila S., & Martín-Loeches, M. (2021). Situating Language in a Minimal Social Context: How Seeing a Picture of the Speaker’s Face Affects Language Comprehension. *Social Cognitive and Affective Neuroscience*. doi: 10.1093/scan/nsab009

Herzmann, G., Schweinberger, S. R., Sommer, W., & Jentzsch, I. (2004). What's special about personally familiar faces? A multimodal approach. Psychophysiology, 41(5), 688-701. doi: 10.1111/j.1469-8986.2004.00196.x

Hommel, B., Proctor, R. W., & Vu, K.-P. L. (2004). A feature-integration account of sequential effects in the Simon task. *Psychological Research*, 68(1), 1-17. doi: 10.1007/s00426-003-0132-y

Hsu, N. S., & Novick, J. M. (2016). Dynamic engagement of cognitive control modulates recovery from misinterpretation during real-time language processing. *Psychological science*, 27(4), 572-582. doi: 10.1177/0956797615625223

Jacques, C., & Rossion, B. (2006). The speed of individual face categorization. Psychological science, 17(6), 485-492. doi: 10.1111/j.1467-9280.2006.01733.x

Jacques, C., d'Arripe, O., & Rossion, B. (2007). The time course of the inversion effect during individual face discrimination. Journal of Vision, 7(8), 3-3. doi: 10.1167/7.8.3

January, D., Trueswell, J. C., & Thompson-Schill, S. L. (2009). Co-localization of Stroop and syntactic ambiguity resolution in Broca's area: Implications for the neural basis of sentence processing. *Journal of Cognitive Neuroscience,* 21(12), 2434-2444. doi: 10.1162%2Fjocn.2008.21179

Johnson, R. (1986). A triarchic model of P300 amplitude. *Psychophysiology*

Jost, K., Mayr, U., & Rösler, F. (2008). Is task switching nothing but cue priming? Evidence from ERPs. *Cognitive, Affective, & Behavioral Neuroscience*, *8*(1), 74-84. doi: 10.3758/CABN.8.1.74

Kaan, E., & Swaab, T. Y. (2003). Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of Cognitive Neuroscience*, 15(1), 98-110. doi: 10.1162/089892903321107855

Kaganovich, N., Francis, A. L., & Melara, R. D. (2006). Electrophysiological evidence for early interaction between talker and linguistic information during speech perception. *Brain research*, 1114(1), 161-172. doi: 10.1016/j.brainres.2006.07.049

Kan, I. P., Teubner-Rhodes, S., Drummey, A. B., Nutile, L., Krupa, L., & Novick, J. M. (2013). To adapt or not to adapt: The question of domain-general cognitive control. *Cognition*, 129(3), 637-651. doi: 10.1016/j.cognition.2013.09.001

Kapadia, A. M., & Perrachione, T. K. (2020). Selecting among competing models of talker adaptation: Attention, cognition, and memory in speech processing efficiency. *Cognition*, 204, 104393. doi: 10.1016/j.cognition.2020.104393

Kerns, J. G., Cohen, J. D., MacDonald, A. W., Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science*, 303(5660), 1023-1026. doi: 10.1126/science.1089910

Key-DeLyria, S. E., & Altmann, L. J. (2016). Executive function and ambiguous sentence comprehension. *American journal of speech-language pathology*, 25(2), 252-267. doi: 10.1044/2015_AJSLP-14-0111

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching — A review. *Psychological bulletin*, *136*(5), 849. doi: 10.1037/a0019842

Koester, D., Gunter, T. C., Wagner, S., & Friederici, A. D. (2004). Morphosyntax, prosody, and linking elements: The auditory processing of German nominal compounds. Journal of Cognitive Neuroscience, 16(9), 1647-1668

Kolk, H. H., Chwilla, D. J., van Herten, M., & Oor, P. J. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. Brain and language, 85(1), 1-36. doi: 10.1016/s0093-934x(02)00548-5

Kolk, H., & Chwilla, D. (2007). Late positivities in unusual situations. *Brain and language*, 100(3), 257-261. doi:10.1016/j.bandl.2006.07.006

Kopp, B., Steinke, A., & Visalli, A. (2020). Cognitive flexibility and N2/P3 event-related brain potentials. *Scientific reports*, *10*(1), 1-17.doi: 10.1038/s41598-020-66781-5

Kroczek, L. O. H., & Gunter, T. C. (2017). Communicative predictions can overrule linguistic priors. *Sci Rep*, 7(1), 17581. doi:10.1038/s41598-017-17907-9

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. Brain research, 1146, 23-49

Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. Cognitive Brain Research, 17(1), 117-129. doi: 10.1016/s0926-6410(03)00086-7

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual review of psychology*, 62, 621-647. doi: 10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. Science, 207(4427), 203-205. doi: 10.1126/science.7350657

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. Cognition and emotion, 24(8), 1377-1388

Lattner, S., & Friederici, A. D. (2003). Talker's voice and gender stereotype in human auditory sentence processing–evidence from event-related brain potentials. *Neuroscience letters*, 339(3), 191-194. doi: 10.1016/s0304-3940(03)00027-2

Leckey, M., & Federmeier, K. D. (2019). The P3b and P600 (s): Positive contributions to language comprehension. *Psychophysiology*, e13351. doi: 10.1111/psyp.13351

Leuninger, H. (1996). Danke und Tschüss fürs Mitnehmen: gesammelte Versprecher und eine kleine Theorie ihrer Korrekturen [Thanks and bye to go: collected slips of the tongue and a small theory about their corrections]. Munich: Ammann

Leuninger, H. (1999). Reden ist Schweigen, Silber ist Gold: gesammelte Versprecher [Talk is silence, silver is golden: collected slips of the tongue]. Leipzig: Dt. Taschenbuch-Verlag

Li, L., Miller, E. K., & Desimone, R. (1993). The representation of stimulus familiarity in anterior inferior temporal cortex. Journal of neurophysiology, 69(6), 1918-1929. doi: 10.1152/jn.1993.69.6.1918

Lim, S.-J., Shinn-Cunningham, B. G., & Perrachione, T. K. (2019). Effects of talker continuity and speech rate on auditory working memory. *Attention, Perception, & Psychophysics*, 81(4), 1167-1177. doi: 10.3758/s13414-019-01684-w

Lim, S.-J., Tin, J. A., Qu, A., & Perrachione, T. K. (2019). Attentional reorientation explains processing costs associated with talker variability. Paper presented at the 19th International Congress of Phonetic Sciences. retrieved from http://sites.bu.edu/cnrlab/files/2019/05/Lim-Tin-Qu-Perrachione-2019-ICPhS-TCTA.pdf

Lindenberger, U., Ebner, N., & Riediger, M. (2005-2007). FACES collection, from Max Planck Institute for Human Development, Center for Lifespan Psychology, Berlin, Germany website: http://faces.mpib-berlin.mpg.de/album/escidoc:57488

Magnuson, J. S., & Nusbaum, H. C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology: Human Perception and Performance*, 33(2), 391. doi: 10.1037/0096-1523.33.2.391

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. Journal of neuroscience methods, 164(1), 177-190

Martín-Loeches, M., Nigbur, R., Casado, P., Hohlfeld, A., & Sommer, W. (2006). Semantics prevalence over syntax during sentence processing: A brain potential study of noun-adjective agreement in Spanish. Brain research, 1093(1), 178-189

Meringer, R., & Mayer, K. (1895). Versprechen und Verlesen: eine Psychologisch-linguistische Studie [Slips of the tongue and Misreading: a psychological linguistic study]. Stuttgart: Göschensche Verlagsbuchhandlung

Miller, E. K., & Desimone, R. (1994). Parallel neuronal mechanisms for short-term memory. Science, 263(5146). doi: 520-522. 10.1126/science.8290960

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. Cognitive Psychology, 41(1), 49-100. doi: 10.1006/cogp.1999.0734

Molinaro, N., Barber, H. A., & Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. Cortex, 47(8), 908-930

Molinaro, N., Vespignani, F., & Job, R. (2008). A deeper reanalysis of a superficial feature: An ERP study on agreement violations. Brain research, 1228, 161-176

Mueller, J. L., Oberecker, R., & Friederici, A. D. (2009). Syntactic learning by mere exposure-an ERP study in adult learners. BMC neuroscience, 10(1), 89

Munro, M. J. (2003). A Primer on Accent Discrimination in the Canadian Context. *TESL Canada journal*, 20(2), 38-51. doi: 10.18806/tesl.v20i2.947

Münte, T. F., Heinze, H.-J., Matzke, M., Wieringa, B. M., & Johannes, S. (1998). Brain potentials and syntactic violations revisited: No evidence for specificity of the syntactic positive shift. *Neuropsychologia*, 36(3), 217-226. doi: 10.1016/S0028-3932(97)00119-X

Novick, J. M., Hussey, E., Teubner-Rhodes, S., Harbison, J. I., & Bunting, M. F. (2014). Clearing the garden-path: Improving sentence processing through cognitive control training. *Language, Cognition and Neuroscience*, 29(2), 186-217. doi: 10.1080/01690965.2012.758297

Nusbaum, H. C., & Morin, T. M. (1992). Paying attention to differences among talkers. In Y. Tohkura, E. Vatikiotis-Bateson, Y. Sagisaka (Ed.), Speech perception, production and linguistic structure (pp. 113-134). Tokyo: Ohmsha, Ltd

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia, 9(1), 97-113. doi: 10.1016/0028-3932(71)90067-4

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational intelligence and neuroscience*, 2011, 1. doi: 10.1155/2011/156869

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, 31(6), 785-806. doi: 10.1016/0749-596X(92)90039-Z

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. Behav Brain Sci, 36(4), 329-347

Rabovsky, M., Hansen, S. S., & McClelland, J. L. (2018). I like coffee with cream and dog? Change in an implicit probabilistic representation captures meaning processing in the brain. bioRxiv, 138-149

Regel, S., Coulson, S., & Gunter, T. C. (2010). The communicative style of a speaker can affect language comprehension? ERP evidence from the comprehension of irony. *Brain research*, 1311, 121-135. doi: /10.1016/j.brainres.2009.10.077

Roehm, D., Bornkessel, I., Haider, H., & Schlesewsky, M. (2005). When case meets agreement: event-related potential effects for morphology-based conflict resolution in human language comprehension. NeuroReport, 16(8), 875-878

Romero-Rivas, C., Martin, C. D., & Costa, A. (2015). Processing changes when listening to foreign-accented speech. *Front Hum Neurosci*, 9, 167. doi:10.3389/fnhum.2015.00167

Rossi, S., Gugler, M. F., Hahne, A., & Friederici, A. D. (2005). When word category information encounters morphosyntax: An ERP study. Neuroscience letters, 384(3), 228-233

Rossion, B. (2014). Understanding face perception by means of human electrophysiology. Trends in cognitive sciences, 18(6), 310-318. doi: 10.1016/j.tics.2014.02.013

Rossion, B., & Jacques, C. (2011). The N170: understanding the time-course of face perception in the human brain. In S. J. Luck, & E. S. Kappenman (Eds.), The Oxford handbook of ERP components (115-142). Oxford: Oxford University Press

Rünger, D., Schwager, S., & Frensch, P. A. (2010). Across-task conflict regulation: A replication failure. *Journal of Experimental Psychology: Human Perception and Performance*, 36(1), 136-146. doi: 10.1037/a0017172

Sabourin, L., Stowe, L. A., & De Haan, G. J. (2006). Transfer effects in learning a second language grammatical gender system. *Second Language Research*, 22(1), 1-29. doi: 10.1191/0267658306sr259oa

Sassenhagen, J., & Fiebach, C. J. (2019). Finding the P3 in the P600: Decoding shared neural mechanisms of responses to syntactic violations and oddball targets. *NeuroImage*, *200*, 425-436. doi: 10.1016/j.neuroimage.2019.06.048

Sassenhagen, J., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2014). The P600-as-P3 hypothesis revisited: Single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned. *Brain and language*, 137, 29-39. doi: 10.1016/j.bandl.2014.07.010

Schacht, A., Sommer, W., Shmuilovich, O., Martienz, P. C., & Martin-Loeches, M. (2014). Differential task effects on N400 and P600 elicited by semantic and syntactic violations. PLoS One, 9(3), e91226. doi: 10.1371/journal.pone.0091226

Schweinberger, S. R., & Neumann, M. F. (2016). Repetition effects in human ERPs to faces. Cortex, 80, 141-153. doi: 10.1016/j.cortex.2015.11.001

Schweinberger, S. R., Pfütze, E.-M., & Sommer, W. (1995). Repetition priming and associative priming of face recognition: Evidence from event-related potentials. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21(3), 722. doi: 10.1037/0278-7393.21.3.722

Schweinberger, S. R., Pickering, E. C., Jentzsch, I., Burton, A. M., & Kaufmann, J. M. (2002). Event-related brain potential evidence for a response of inferior temporal cortex to familiar face repetitions. Cognitive Brain Research, 14(3), 398-409. doi: 10.1016/s0926-6410(02)00142-8

Shinn-Cunningham, B.G. 2008. Object-based auditory and visual attention. *Trends Cogn. Sci*. 12, 182-186. doi: 10.1016/j.tics.2008.02.003

Simon, J. R. (1969). Reactions toward the source of stimulation. *Journal of experimental psychology*, *81*(1), 174. doi: 10.1037/h0027448

Sommer, W., Leuthold, H., & Matt, J. (1998). The expectancies that govern the P300 amplitude are mostly automatic and unconscious. *Behavioral and Brain Sciences*, *21*(1), 149-150. doi: 10.1017/S0140525X98210958

Sommer, W., Matt, J., & Leuthold, H. (1990). Consciousness of attention and expectancy as reflected in event-related potentials and reaction times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(5), 902. doi: 10.1037/0278-7393.16.5.902

Stürmer, B., Leuthold, H., Soetens, E., Schröter, H., & Sommer, W. (2002). Control over location-based response activation in the Simon task: behavioral and electrophysiological evidence. *Journal of Experimental Psychology: Human Perception and Performance*, 28(6), 1345. doi: 10.1037/0096-1523.28.6.1345

Thothathiri, M., Asaro, C. T., Hsu, N. S., & Novick, J. M. (2018). Who did what? A causal role for cognitive control in thematic role assignment during sentence comprehension. *Cognition*, 178, 162-177. doi: 10.1080/02687038.2017.1350630

Ullsperger, M., Bylsma, L. M., & Botvinick, M. M. (2005). The conflict adaptation effect: It's not just priming. *Cognitive, Affective, & Behavioral Neuroscience*, 5(4), 467-472. doi: 10.3758/CABN.5.4.467

van Berkum, Jos. (2004). Sentence comprehension in a wider discourse: Can we use ERPs to keep track of things?. In M. Carreiras & C. Clifton, Jr. (2004) (Eds.), The on-line study of sentence comprehension: Eyetracking, ERPs and beyond (pp. 229-270). New York: Psychology Press

van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. Journal of Experimental Psychology: Learning, Memory, and Cognition, 31(3), 443. doi: 10.1037/0278-7393.31.3.443

van Berkum, J. J., van den Brink, D., Tesink, C. M., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, 20(4), 580-591. doi: 10.1162/jocn.2008.20054

van de Meerendonk, N., Indefrey, P., Chwilla, D. J., & Kolk, H. H. (2011). Monitoring in language perception: Electrophysiological and hemodynamic responses to spelling violations. *Neuroimage*, 54(3), 2350-2363. doi: 10.1016/j.neuroimage.2010.10.022

van de Meerendonk, N., Kolk, H. H., Chwilla, D. J., & Vissers, C. T. W. (2009). Monitoring in language perception. *Language and Linguistics Compass*, 3(5), 1211-1224. doi: 10.1111/j.1749-818X.2009.00163.x

van den Brink, D., Brown, C. M., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. Journal of Cognitive Neuroscience, 13(7), 967-985 van Herten, M., Kolk, H. H., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Brain Res Cogn Brain Res*, 22(2), 241-255. doi:10.1016/j.cogbrainres.2004.09.002

van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. Journal of Experimental Psychology: Learning, Memory, and Cognition, 25(2), 394

Vissers, C. T. W., Kolk, H. H., van de Meerendonk, N., & Chwilla, D. J. (2008). Monitoring in language perception: evidence from ERPs in a picture–sentence matching task. *Neuropsychologia*, 46(4), 967-982. doi: 10.1016/j.neuropsychologia.2007.11.027

Vissers, C. T., Chwilla, D. J., & Kolk, H. H. (2006). Monitoring in language perception: The effect of misspellings of words in highly constrained sentences. Brain Res, 1106(1), 150-163. doi: 10.1016/j.brainres.2006.05.012

von Kriegstein, K., Dogan, O., Grüter, M., Giraud, A.-L., Kell, C. A., Grüter, T., Kleinschmidt, A., Kiebel, S. J. (2008). Simulation of talking faces in the human brain improves auditory speech recognition. Proceedings of the National Academy of Sciences, 105(18), 6747-6752. doi: 10.1073/pnas.0710826105

Vuong, L. C., & Martin, R. C. (2014). Domain-specific executive control and the revision of misinterpretations in sentence comprehension. *Language, Cognition and Neuroscience*, 29(3), 312-325. doi: 10.1080/01690965.2013.836231

Weissgerber, T. L., Milic, N. M., Winham, S. J., & Garovic, V. D. (2015). Beyond bar and line graphs: time for a new data presentation paradigm. PLoS Biol, 13(4), e1002128. doi: 10.1371/journal.pbio.1002128

Wiese, H., Tüttenberg, S. C., Ingram, B. T., Chan, C. Y., Gurbuz, Z., Burton, A. M., & Young, A. W. (2019). A robust neural index of high face familiarity. Psychological science, 30(2), 261-272. doi: 10.1177/0956797618813572

Wietzel-Winkler, S. (2017, May). Unsere Versprecher & Verhaspler [Our slips of the tongue]. Retrieved from http://www.zur-sprache-kommen.de/?Unsere_Versprecher

Wiggs, C. L., & Martin, A. (1998). Properties and mechanisms of perceptual priming. Current opinion in neurobiology, 8(2), 227-233. doi: 10.1016/s0959-4388(98)80144-x

Wong, P. C., Nusbaum, H. C., & Small, S. L. (2004). Neural bases of talker normalization. *Journal of Cognitive Neuroscience*, 16(7), 1173-1184. doi: 10.1162/0898929041920522

Xu, J., Abdel Rahman, R., & Sommer, W. (2019). Perceived language competence modulates criteria for speech error processing: evidence from event-related potentials. *Language, Cognition and Neuroscience*, 1-14. doi: 10.1080/23273798.2018.1562558

Xu, J., Abdel Rahman, R., & Sommer, W. (2021). Sequential Adaptation Effects Reveal Proactive Control in Processing Spoken Sentences: Evidence from Event-related Potentials. *Brain and Language*, 214, 104904. doi: 10.1016/j.bandl.2020.104904

Ye, Z., & Zhou, X. (2009). Executive control in language processing. *Neuroscience & Biobehavioral Reviews*, 33(8), 1168-1177. doi: 10.1016/j.neubiorev.2009.03.003

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt,

1. dass keine Zusammenarbeit mit gewerblichen Promotionsberatern stattfand,

2. dass ich die Dissertation auf der Grundlage der angegebenen Hilfsmittel und Hilfen selbstständig angefertigt habe,

3. dass ich mich nicht anderwärts um einen Doktorgrad beworben habe bzw. einen entsprechenden Doktorgrad besitze,

4. dass mir die dem angestrebten Verfahren zugrunde liegende Promotionsordnung der Lebenswissenschaftlichen Fakultät vom 05. März 2015, veröffentlicht im Amtlichen Mitteilungsblatt der Humboldt-Universität zu Berlin Nr. 12/2015 bekannt ist,

5. dass die Dissertation oder Teile davon nicht bereits bei einer anderen wissenschaftlichen Einrichtung eingereicht, angenommen oder abgelehnt wurden,

6. dass die Grundsätze der Humboldt-Universität zu Berlin zur Sicherung guter wissenschaftlicher Praxis eingehalten wurden.

Berlin, den 16. Februar 2021

Jue Xu